STRUCTURE AND FUNCTION OF THE DELETED IN AZOOSPERMIA GENE

A Dissertation

by

DAVID CHASE CAMERON SPRAGUE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: Medical Sciences

STRUCTURE AND FUNCTION OF THE DELETED IN AZOOSPERMIA GENE

A Dissertation

by

DAVID CHASE CAMERON SPRAGUE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Thomas J. Kuehl
	Allison C. Rice-Ficht
Committee Members,	Fuller W. Bazer
	Steve A. Maxwell
Head of Department,	J. Martin Scholtz

December 2006

Major Subject: Medical Sciences

ABSTRACT

Structure and Function of the Deleted in Azoospermia Gene. (December 2006)

David Chase Cameron Sprague, B.S., Saint Edward's University

Co-Chairs of Advisory Committee: Dr. Thomas J. Kuehl
Dr. Allison C. Rice-Ficht

A number of genes have been associated with variation in human spermatogenesis related to fertility. One of these, the Deleted in Azoospermia (DAZ) gene, exists as copies on two chromosomes, 3 and Y. The autosomal copy, DAZ-like (DAZL), has one RNA recognition motif (RRM) and is homologous to the DAZL gene found throughout the vertebrate lineage. There are four copies of DAZ on the Y chromosome with a pair at each of two sites. One pair contains a single RRM and the other has three RRMs. Human DAZ is homologous to genes in old world primates and ape Y chromosomes. Both DAZ and DAZL bind messenger RNAs at U-rich sequences near the poly-A tail in a manner that facilitates translation. Both are expressed in spermatogonia during the transition from mitotic cellular expansion through meiotic chromosomal reduction and during spermiogenesis. This study examined genomic variation in DAZ and DAZL, including deletion of DAZ from individuals with various levels of sperm cell production and mutations of DAZL in male partners of infertile couples. Deletions in DAZ are not as common in azoospermic men from central Texas as compared to other reports. Single nucleotide polymorphisms (SNPs) were identified in anonymous infertility patients, but were not located in the exons of the RRM. Proteins produced from transcripts encoded by genes from human DAZL, DAZL with

SNPs within and outside the RRM, and a DAZ with single RRM were identified. Binding activity of DAZL to mRNA was confirmed using a microarray method, and mRNA from human testes was screened to identify at least 1,313 mRNA potential targets for DAZL. These targets were involved in ribosome construction, pyruvate metabolism, cell cycle control, and proteasome function. Variations in binding of protein to a high and a low bound target mRNA were demonstrated between protein constructs of DAZL, DAZL with mutations, and DAZ. Binding of DAZL to mRNA was also confirmed using electrophoretic mobility shift assays. With materials and procedures developed during this study, comparisons of genetic variants of DAZ and DAZL can be performed to identify mechanisms responsible for structural and functional differences in control of spermatogenesis.

DEDICATION

To my Wife, Dawn.

There is no other way to say it without being trite, but I couldn't have done it without you. You have been with me for the whole way. You gave me encouragement when I needed it most and comforting when I could not go on any further. Mostly, you reminded me to forgive when I got lost in myself.

Thank you and I love you very much. …and this is for you too- William, and Maren.

ACKNOWLEDGEMENTS

I would like to thank both of my co-principal investigators, Dr. Thomas J. Kuehl and Allison C. Rice-Ficht, for their continual support of my project, my professional growth, and my personal development. I have learned much in addition to the book work and lab work. Additionally, Dr. Bazer and Dr. Maxwell used their valuable time to review my work over many iterations and for that I am grateful.

I also want to thank the faculty in the Department of Molecular & Cellular Medicine for guidance. There also have been numerous staff at Scott & White over the years who have contributed greatly to this project and include, Monica Brown, Patrick Conley, Janet Dye, and Rob LeFever. My peers during my tenure at Texas A&M, who supported me include Kenny Carson, Brian O'Shea, Angela Arenas, Barbara Ruef, Mike Remedi, Jason Schmittschmitt, and Sean Conlan. Finally, I thank the support staff in my department- Lydia Mousner, Chantel Plaag, Janis Chmiel, Rebecca Hogard, and J.D. Luza.

I also received a great amount of support from my family. Thank you Da, Cha, D, Quin, Momma, Jim, and Nancy. Thank you for your patience with me!

Especially, I would like to honor my older brother Quin whose lifelong interest in the sciences served as a starting point for my own scientific endeavors.

You all are number one in my book.

TABLE OF CONTENTS

CHAPTER Page

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION: INFERTILITY AND THE DAZ FAMILY OF GENES

Infertility

Human infertility is associated with male factors in about 50% of cases[1]. Male infertility has been estimated at 15% in Western populations and can be explained by problems with hormones, immunology, sperm delivery, and the physical structure of the reproductive organs [2]. However, in a number of cases, the cause of infertility is unknown due to the lack of knowledge of genetic and molecular mechanisms which play a role in the control of sperm cell production and maturation. One assessment for fertility in males is based on sperm concentration in the semen[3]. Azoospermia indicates no sperm cells in the semen while severe oligozoospermia is the presence of up to five million spermatozoa per milliliter in the ejaculate. Oligozoospermia is the presence of between five and twenty million sperm cells per milliliter in the semen and normozoospermia is the presence of twenty or more million sperm cells per milliliter. To briefly review, spermatogenesis begins with the primordial germ cells (PGCs) increasing in number through multiple rounds of mitosis[1]. The PGCs then migrate to the developing gonadal tissue where they proceed to divide further through mitosis to maintain their numbers or go on to a pathway that leads to meiosis and spermatogenesis.

This thesis follows the style of Biology of Reproduction.

The pathway to spermatogenesis begins with the division to produce spermatogonial cells that enter into meiosis. Meiosis results in haploid germ cells where the final product is fully functional mature spermatozoa. Producing these cells properly has been attributed to more than 3,000 genes operating in conjunction with one another through experimentation involving knockouts in the mouse and fly[4]. Defects in any of these genes has the potential to affect germ cell development. Genetic abnormalities are involved in about 30% of all male infertility, many of which manifest themselves in the form of genetic mutations affecting spermatogenesis[4]. The other 70% of male infertility can be explained by defects in anatomy, errors within the complex myriad of endocrinological pathways, problems with the immune system, and the environmental insults. Genomic errors can generally be classified into three different groups[4]. The first group contains aneuploidies which can be found in 14% of azoospermic patients and 3% of oliogozoospermic patients. The largest group of patients with aneuploidy is the group containing Kleinfelter patients (47,XXY). The majority of these patients are azoospermic, but the exact cause of the infertility is not currently known. The second group contains the presence of submicroscopic deletions or "microdeletions" which are deletions of groups of genes on the Y chromosome involved in spermatogenesis as well as rearrangement of those genes (examples include duplications, inversions, and amplifications). The term microdeletion was coined to describe deletions on the Y chromosome that were not detectable to whole chromosomal staining[4]. Thirdly, there also is the altered expression of a single gene that leads to infertility of which an example is a point mutation leading to infertility which is found within the Androgen

Insensitivity Syndromes (AIS)[4]. AIS include testicular feminization, Reifenstien, and Infertile Male syndrome where a point mutation in the androgen receptor (AR) gene leads to infertility. Androgen receptor binds either testosterone or dihydrotestosterone, is activated, dimerizes, is transported to the nucleus, and binds DNA. However, point mutations in either the transactivation domain or the androgen binding domain reduce the ability of the downstream genes to organize sexual differentiation.

Structure of Genomic DNA on the Y Chromosome

Deletions of the Y chromosome were first associated with infertility by Teipolo and Zuffardi[5]. The area specifically deleted was the q11 band of the Y chromosome and has since been termed the azoospermia factor (AZF) region[4]. In 1992, to search for a

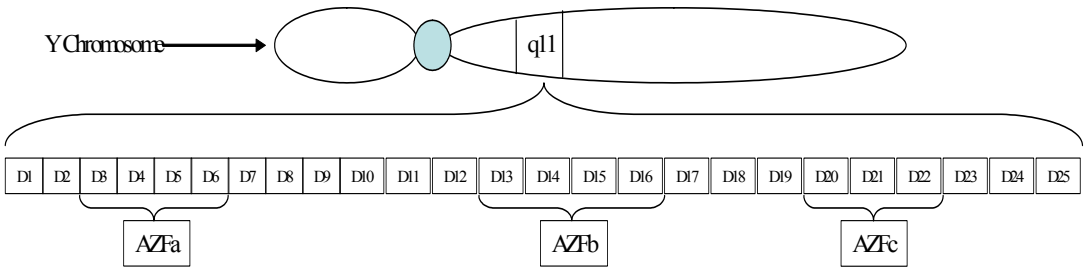


Figure 1. Y Chromosome, Deletion Intervals, and AZF Regions. Adapted from Vogt et al. 2004. The sizes of each of the deletion intervals are arbitrary sizes and no conclusions can be made with regard to comparison between each of the intervals. However, the size of genomic DNA which includes AZFb and AZFc is about 8 Mb in size.

gene or set of genes deleted on the Y chromosome of infertile patients, Vollrath et al. created a deletion map of the Y chromosome from 96 males with partial deletions which is highlighted in Figure 1[6]. Yeast artificial chromosomes containing DNA fragments of the patients were sequenced, aligned, and ordered to give a complete, contiguous sequence. Deletion intervals were identified as areas where patients were deleted on the Y chromosome as compared with normal patients. Where common deletions were discovered, an interval was created in association with a particular genotype. Deletion intervals were defined using PCR primers and the sequences were used to further define the regions of AZFa, AZFb, and AZFc. Since the sequence of the primers was known as well as what comprised the AZF regions, it could be deduced which deletion intervals contained each of the AZF regions. In 1993, Ma et al. suggested that a gene family residing in the AZF region of the Y chromosome contained RNA-binding motifs[7]. The AZF region has been further subdivided into three separate regions termed AZFa, AZFb, and AZFc; each involved in different presentations of male infertility as illustrated in Figure 1. The phenotype in patients with deletions in AZFa is described as "Sertoli Cell Only" which means that there are no spermatids or spermatogonia present in the seminiferous tubules[8]. Sertoli cells function to support the developing sperm during the early stages of spermatogenesis. AZFb deletions are associated with sperm cell maturation stalling at the spermatocyte stage where the cells appear to have normal cytology, but no postmeiotic cells are present. The main outcome of a complete deletion in the AZFc region is severe infertility[9]. AZF deletions, specifically in the distal Yq11 region, result in variability in the development of spermatozoa. Some tubules only have

Sertoli cells while others contain sperm cells at various stages of meiosis and development. Partial deletions in the AZF region are responsible for removing some of the DAZ genes on the Y chromosome which reside within the AZFc region. To date, there is not complete agreement as to whether partial deletions of the DAZ genes afford reduced fertility to the same extent as complete deletion of the AZFc region [9-11].

Within the AZFc region, a sequence-tagged site (STS), sY254, is the most commonly deleted fragment of DNA in patients with reduced fertility[12]. Sequence-tagged sites are unique pieces of DNA that can be detected by the polymerase chain reaction (PCR) and serve as "landmarks" for genome mapping procedures[6]. The sY254 resides within the RNA-binding domain of a gene termed Deleted in Azoospermia, or DAZ[12] that was identified through nucleotide analysis of the gene from a cDNA library derived from human testis. However, at the time it was thought that there was only a single copy of the gene represented by sY254. Subsequent research on AZFc indicated the presence of four copies of the DAZ gene and a highly variable number of DAZ copies [13, 14].

DAZ Family of Proteins

Figure 2 illustrates DAZ gene configurations found on the Y chromosome and the autosomal homologue of DAZ, termed DAZL for DAZ-like. All family members consist of at least one RRM (from one to three) at the N terminus and a 24 amino acid motif repeated 9 to15 times at the C-terminus (DAZ-repeat domain). Researchers in Taiwan described a cohort of patients with reduced fertility whose RRMs in DAZL

contained a point mutation within the RRM and another point mutation outside of the RRM[15]. These mutations are indicated as vertical yellow and red bars in Figure 2. In addition, the structure of the DAZ genes is further complicated by the fact that different patients exhibit polymorphisms in the number of repeats in the repeat region of the DAZ gene as well as the composition of the repeats themselves within the region[16]. For example, patient A may have fifteen repeats in DAZ2 as indicated in Figure 2. However, patient B may have fourteen repeats in DAZ2 and also have a different complement of repeats B, C, D, E, F, G,H, and I (Table 1). Furthermore, the repeat region is currently thought to be the facilitator of protein-protein interactions between DAZ and other proteins [17]. When the RRMs of human and mouse are compared it becomes evident that there are no differences in the amino acid sequence (Figure 3). On the other hand, the DAZL RRM and the DAZ RRM differ in a few places. At position seven of the RRM, DAZL contains a methionine whereas DAZ contains a valine. At position 18, DAZL contains a valine while DAZ contains an alanine residue. At position 29, DAZL contains phenylalanine where DAZ contains a cytosine and so forth. Of note is that none of the differences between DAZ and DAZL lie within the two ribonucleoprotein (RNP) domains. RNP1 and RNP2 are conserved sequences which describe RRM domains[18]. RNP1 is well conserved as an octapeptide containing the sequence Lys/Arg – Gly – Phe/Tyr – Gly/Ala – Phe – Val – X – Phe/Tyr[19]. The RNP2 hexapeptide is less well conserved but is rich in aliphatic residues such as valine, glycine, and isoleucine as well as aromatic residues such as phenylalanine.

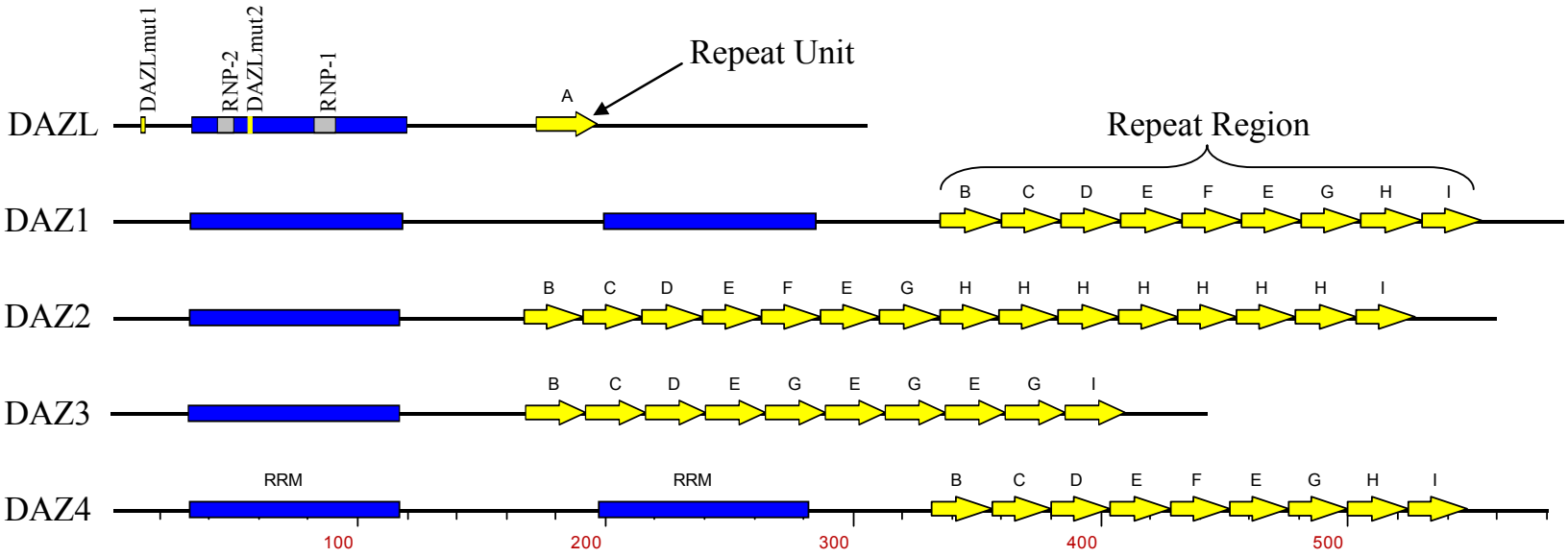


Figure 2. DAZ Family Gene Structure. Blue boxes indicate the RRMs . Small yellow triangles indicate the repeat region and the letters within the triangles indicate different repeats based on sequence. Within the DAZL RRM, DAZLmut2 shows the mutation from the Taiwanese public resulting in a correlation with reduced fertility. Just outside of the RRM is another line (DAZLmut1) which indicates a mutation in the Taiwanese public that is not associated with reduced fertility. In addition, the RNP domains are indicated. Sequences were obtained from Genbank and the graphics produced with Accelrys.

Table 1. DAZ Repeat Sequences.

Repeat	Sequence
A	AYPTYPNSPVQVITGYQLPVYNYQ
B	AYSAYPHSPGQVITGCQLLVYNYQ
C	EYPTYPDSAFQVTTGYQLPVYNYQ
D	PFPAYPRSPFQVTAGYQLPVYNYQ
E	AFPAYPNSPFQVATGYQFPVYNYQ
F	PFPAYPSSPFQVTAGYQLPVYNYQ
G	AFPAYPNSPVQVTTGYQLPVYNYQ
H	AFPAYPSSPFQVTTGYQLPVYNYQ
I	AFPAYPNSAVQVTTGYQFHVYNYQ

The left column indicates the lettered repeat found in Figure 2. The right-hand column indicates the sequence for each of the repeats.

Figure 3 illustrates that both the DAZ and DAZL RRMs fit both of these criteria. At least for the U1 spliceosomal protein, the RRM domain folds into an alternating series of alpha helices and beta sheets with the RNP domains next to each other on opposite beta sheets in an antiparallel fashion[20]. Noticing this feature, Ruggiu et al., 2000 proceeded to knock out the functionality of the mouse Dazl protein by modifying its RRM[21]. Their experiments demonstrated that the RNP domains were required for RNA binding.

```
1- LPEGKIMPNTVFVGGIDVRMDETEIRSFFARYGSVKEVKIITDRTGVSKGYGFVSFFNDVDVQKIVESQINFHGKKLKLGPAIR
2- LPEGKIMPNTVFVGGIDVRMDETEIRSFFARYGSVKEVKIITDRTGVSKGYGFVSFYNDVDVQKIVESQINFHGKKLKLGPAIR
3- LPEGKIVPNTVFVGGIDARMDETEIGSCFGRYGSVKEVKIITNRTGVSKGYGFVSFVNDVDVQKIVGSQIHFHGKKLKLGPAIR
4- LPEGKIMPNTVFVGGIDVRMDEAEIRSFFARYGSVKEVKIITDRTGVSKGYGFVSFFNDVDVQKIVESQINFHGKKLKLGPAIR
5- LPEGKIVPNTVFVGGIDARMDETEIGSCFGRYGSVKEV
6- LPEGKIMPNTVDVGGIDVRMDETEIRSFFARYGSVKEVKIITDRTGVSKGDGDVSDYNDVDVQKIVESQINFHGKKLKLGPAIR
             RNP-2                                    RNP-1
```

Figure 3. Comparison of RRMs of Mouse and Variants Found in Humans. Line 1 is the human DAZL RRM, Line 2 is the mouse Dazl RRM, Line 3 is the human DAZ RRM, Line 4 is the RRM containing DAZLmut2 which has been associated with infertility in the Tiawanese public. Line 5 is sY254's complement of the RRM, and Line 6 is the mouse knockout RRM. Line 5 shows where sY254 would amplify in genomic DNA and Line 6 contains the RRM from a mouse that was blocked of Dazl function with mutations within the RNP domains which are indicated by underlining. The two underlined RNP (ribonucleoprotein) domains are required features of RNA-binding proteins.

An autosomal copy of DAZ, DAZL, found on chromosome 3 has been described as a, "…master control gene for premeiotic germ cell development," functioning as an RNA-binding protein expressed in the germline of both males and females[4]. DAZL is conserved in vertebrates (mouse, zebrafish, and *xenopus*) and in invertebrates (*c. elegans* and *drosophila*)[22]. DAZL protein is expressed in the tails of spermatozoa and its RNA transcript is also found in the residual cytoplasm of mature spermatozoa[23]. The function of Dazl, the autosomal homologue of DAZ in mice, indicates a role in both spermatogenesis and oogenesis since disruption of the gene leads to sterility in both males and females [24]. Much of the information on DAZ function was derived from the *Drosophila* system in which the DAZ homologue, *boule*, is required for spermatogenesis; when knocked out, spermatogenesis is arrested at the G2/M phase of meiosis [25]. The mechanism behind this suggests that *boule* plays a role in translational control of specific mRNAs required for cell cycle progression from G2 to M during mitosis [26]. Delayed translation of specific subsets of mRNA is a well documented

regulatory mechanism of key importance in spermatogenesis [27]; [28, 29]. As an example, in the mouse model, Prbp (Prm-1 RNA binding protein) binds to the 3'UTR of Prm-1 mRNA to repress translation for several days during the round spermatid stage of spermatogenesis. Prm1 (mouse protamine 1) is a protein involved in replacing histones as packaging constituents of DNA and stabilizes DNA in late spermatogenesis. Importantly, regulation of signals in the developing gametes is primarily by post-transcriptional means and is necessary as one major goal during development in the packaging of genetic material into a dense, transcriptionally-silent structure[1]. The mRNAs are produced early in spermatogenesis, genomic DNA becomes silent through condensation with protamines, and protein production is supported through the effects of proteins such as those of the DAZ family.

The DAZ proteins have a variety of interactions including homodimerization and interaction with other proteins expressed during spermatogenesis[17, 21]. Using the yeast two-hybrid assay, DAZ was found to interact with two other proteins, DAZAP1 and DAZAP2 (DAZ Associated Protein) [17]. DAZAP1 is a gene whose product is an RNA-binding protein, but DAZAP2 has not been associated with any function. Both of the proteins bind to DAZ and DAZL through the repeat region, whereas DAZL only has one "repeat." Neither of these proteins was linked to disease states such as reduced fertility based on mapping techniques. In addition, in the mouse, Dazl was found to homodimerize *in vivo* and *in vitro* in a manner independent of RNA[21]. Considering these interactions, the expression of specific subsets of family members may well produce widely different outcomes with respect to sperm viability and function.

Early investigations involving potential spermatogenesis proteins in the AZF region included RBMY, BPY2, and CDY1[30]. The RBMY (RNA-binding motif Y) protein contains an RNA-binding domain and is found in the nucleus of germ cells. The BPY2 (basic charge, y-linked 2) protein is not well characterized, but may play a role in spermatogenesis[31]. The CDY1 (chromodomain protein, Y-linked 1) protein possesses a chromodomain and a histone acetyltransferase domain[32]. The chromodomain is implicated in gene repression while the histone acetyltransferase domain is involved in histone hyperacetylation. Histone hyperacetylation facilitates the transition from histone-coated DNA to protamine-coated DNA during DNA packaging for spermatogenesis. In addition, CDY1 is found in the nucleus of late spermatids.

Expression patterns of the DAZ family proteins led researchers to determine that the human DAZ and DAZL genes are initially expressed in primordial germ cells and continue to be expressed throughout meiosis to haploid post-meiotic germ cells[1]. Figure 4 illustrates the expression patterns in different species over the course spermatogenesis. Antibody staining showed that the DAZ, DAZL, and *boule* genes are expressed exclusively in germ cells. Of interest is the comparison between the expression patterns between the mouse and human. Mouse Dazl is expressed from the beginning of mitosis to the beginning of differentiation of the sperm cells. However, in

Organism	Gene	Mitosis	Meiosis	Differentiation
Drosophila	boule			
C. elegans	daz-1			
Zebrafish	zDAZL			
Xenopus	Xdazl			
Mouse	Boule			
	Dazl			
Marmoset	DAZL			
Macaque	DAZ/DAZL			
Human	BOULE			
	DAZL			
	DAZ			

Figure 4. Expression Profiles of the DAZ Family of Genes in Different Organisms. This graphic represents many different experiments by different researchers where the main product detected is protein localization through antibody staining Adapted from Reynolds et al. [1].

humans, the expression of human DAZL is from mitosis until the end of differentiation. Interestingly, DAZ contains the same expression pattern in humans as in mice. This suggests similar, but different roles of DAZ and DAZL in human versus mouse reproduction. Therefore, it may not be possible to fully elucidate the role of the human DAZ and DAZL genes in the mouse model.

Deletions, Inversions, and Reordering of Genomic DNA in the AZF Region

The presence of repeated regions in DAZ genes as well as the presence of multiple gene copies may promote recombination and alterations in gene sequence [33]. Since the DAZ genes on the Y chromosome exist as multiple copies and indeed the DAZ gene itself contains internal repeated motifs, many deletions and rearrangements in the

DAZ region affecting expression are likely to occur. It is also clear that the DAZ region is polymorphic, with individual males exhibiting different copy numbers and arrangements of DAZ genes as well as different numbers of DAZ repeats within DAZ genes [16]. Deletions in the AZF regions occur through recombination events of homologous sequences common in the AZF regions of the Y chromosome [34, 35].

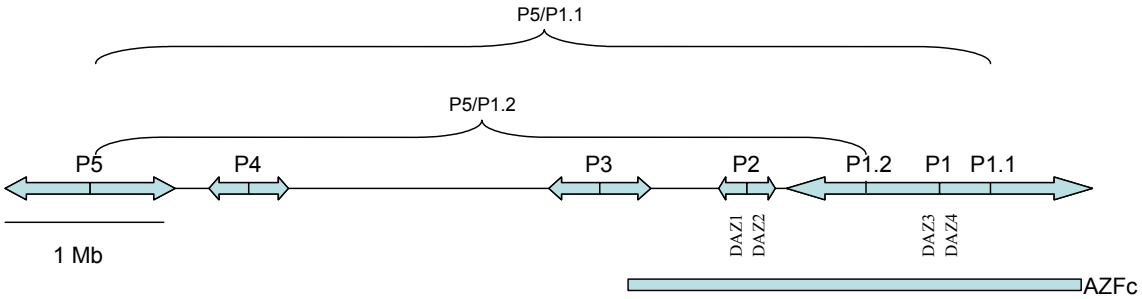


Figure 5. Palindromes in the AZF Region. Palindrome P5 can roughly be used as a size marker of 1 Mb. There are five major palindromes of interest in the AZF region. Within palindrome 1, there are two minipalindromes: P1.2 and P1.1. The AZFc region is defined by the distal portion of palindrome 3 to the distal portion of palindrome 1. Within palindrome 1 are two minipalindromes: palindrome 1.2 and palindrome 1.1. Depending on where homolgous recombination occurs, either all of the DAZ genes are removed or only two of them. Adapted from Repping et al., 2002.

Within the AZFb and AZFc regions, the genomic sequence is arranged in sets of five palindromes (Figure 5)[33]. Because of the complex nature of the palindromic sequences, different outcomes to the deletions can be seen. Figure 5 highlights two of the more common examples of deletions due to the presence of palindromic sequences. For example, if an ectopic recombination event takes place between palindromes 5 and 1.1, the entire AZFc region is deleted along with all copies of the DAZ genes which reside in palindromes 2 and 1. However, if a recombination event occurs between

palindromes 5 and 1.2, then DAZ3 and DAZ4 remain. The research involved in determining what events took place in certain patients involved the use of PCR primers designed specifically to amplify only regions that were not repeated within the AZF region[36]. The PCR products afforded the researchers patterns of plus/minus hits from which the genomic sequence could be deduced and, therefore, the events leading to the deletions determined.

The AZF region is complicated by deletions or duplications arising from sister chromatid exchange between repeated regions within AZF[14]. Through the mechanism highlighted in Figure 6, it is possible to delete all copies of DAZ, or delete two copies, retain four normal copies, gain two copies, or gain four copies. In addition, inversions can occur due to intrachromsomal exchange between inverted repeats found common within the AZF region[37]. When all of the possible mutational events due to the repetitive and polymorphic nature of the AZF region are taken into account, the complexity of the AZF region can be appreciated.

Putative Role for DAZL

The function of DAZ is currently under intense scrutiny by a number of groups. Recently a possible role for DAZL as a recruitment protein for poly-A binding protein (PABP) in Xenopus was introduced[38]. It was found that DAZL bound the 3' untranslated region (3'UTR) on a reporter mRNA construct and recruited PABP to link to the mRNA construct. In translation the first step begins with the recruitment of the cap complex to the 5' methyl-G cap. Next, the small ribosomal subunit binds to the

initiator codon (AUG) at which point the large ribosomal subunit binds. Initiation is then stimulated with protein construction. Multiple ribosomal units can initiate and follow the mRNA producing a structure called a polysome. When each ribosome reaches a stop codon, the subunits separate. Circularization of the mRNA by the poly-A binding proteins binding the poly-A tail and escorting it close to the 5' end of the transcript would allow for increased translation as the ribosomal units in the neighborhood of the initiator codon[38]. The authors suggested that DAZL binds to the 3' UTR of mRNAs with shortened poly-A tails to recruit PABP for stimulation of translation via this circularization mechanism. As DAZL and DAZ share similarities in their RRM, it is possible that these proteins function in similar fashions to augment translation. However, the presence of some variation could relate to specific differences in mRNA specificities or to differences between species so that both DAZL and DAZ are needed to allow normal progression of spermatogenesis in those species with both genes. Some overlap in function between DAZL and DAZ is also likely as patients with deleted DAZ genes may still produce some sperm cells that are present in their ejaculate (Table 2, page 18). While variation in DAZ copy number and the mechanisms that produce this variation have been cited previously in this section, only the absence of all copies of DAZ has been consistently related to the phenotype of male infertility [9-11]. In addition, the role of variation between the separate copies of DAZ, for example those with three RRM regions versus one RRM region, is currently unknown.

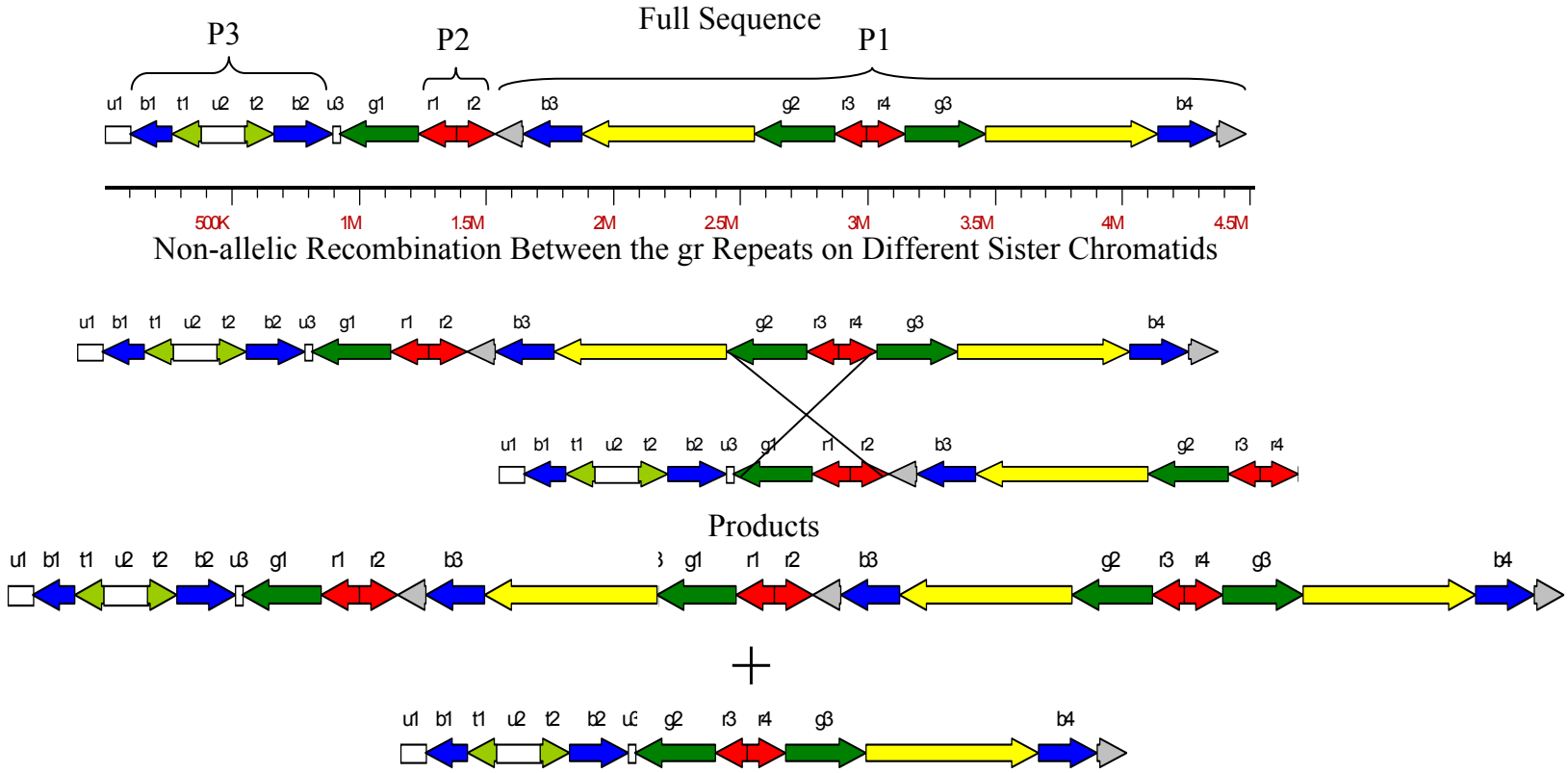


Figure 6. Gene Duplication and Deletion in AZFc by Unequal Crossing Over. The gr/gr deletion shown above is a deletion mechanism shown to increase or decrease the number of DAZ genes shown by the green arrows labeled with r1, r2, r3, and r4. The "r" number refers to the DAZ gene number and is labeled red. Different sister chromatids can have errors in exchange to produce these results of two versions- one with 2 DAZ genes and one with six DAZ genes. Adapted from Lin et al., 2005. The terminology above each of the arrows indicate the color that was assigned to them by researchers from David Page's group. They indicate the color of the arrows and are representative of identical sequences found within the AZF region. Grey arrows indicate unique regions within the AZF region. Additionally, palindromes are indicated above in brackets.

CHAPTER II

GENOMIC ANALYSIS OF DAZ

Introduction

Infertility is a medical condition responsible for frustration and significant expenditures of healthcare dollars affecting about 15% of couples during their reproductive ages (20 to 40 years of age). Infertility can be attributed to male factors in 50% of such couples. The mechanisms of male factor-related causes of infertility act primarily through reducing numbers of functional spermatozoa that are capable of capacitation and fertilization. In such cases, though the male may be the primary affected partner, the most effective treatments involve significant manipulation and risk to the female in the process of obtaining ova for fertilization using assisted reproductive technologies. The combination of the development of techniques for enhancing fertilization using affected sperm cells and advancements in molecular biology of the Y chromosome have resulted in identification of genes involved in gametogenesis[13]. In particular, at least three regions of the Y chromosome contain genes that are occasionally deleted in males with defects in spermatogenesis [39]. Research interests in such genes are two-fold. The first is to understand the frequency of such deletions in male partners of a general population of infertile couples to elucidate the consequences of microdeletions as a form of "inherited infertility." The second is to trace the molecular function of at least one of the genes to improve understanding of genetic regulation of spermatogenesis. Measurements of molecular function may identify

portions of the gene product that are critical and where mutations would limit function. This insight would lead to improvements in diagnostic screening tests that go beyond assessment of the presence or absence of genes to presence of intact, functional gene products using reagents and methods currently unavailable. This study focuses on the DAZ gene (deleted in azoospermia).

While numerous publications, (Table 2), document the frequency of microdeletions of the DAZ region of the Y chromosome in males referred to centers for specialized assisted reproduction procedures, few reports provide an assessment of the frequency in a general population of male partners of infertile couples. The reports for the centers cited in Table 2 suggest frequencies of DAZ deletions on the order of 11% among males with few or no sperm in the ejaculate and on the order of 6% for males with sperm concentrations up to 5 million/ml. As infertility from Y chromosome microdeletions affecting spermatogenesis can be overcome by assisted reproduction techniques when even a few sperm cells are present, the males born to couples using these procedures will inherit their fathers' affected Y chromosome**.** Therefore, knowledge of the frequency of microdeletions is needed to justify decisions about adding DNA screening procedures to routine semen analysis and for couples considering assisted reproduction. The tagged sequence site (sY254) is located within the RNA-binding domain of the DAZ genes[40]. The absence of sY254 has been used as a diagnostic test for the microdeletion of genes which reside in AZFc (Table 2). Research on DAZ has focused on screening for deletions in the Y chromosome in areas where the DAZ genes reside. A possibility exists that the DAZ proteins produced from these genes

might also be defective. Defective proteins might be produced by minute defects in a region critical for protein function.

In addition to large deletions, an attractive working hypothesis suggests that males may not be deleted in DAZ or DAZL, but might possess mutations in the RRM of any or each of the proteins. Mutations in the RRM could lead to a conformational change in DAZ proteins due to amino acid substitutions that affect binding to RNAs. One type of mutation frequently found throughout the genome is the single nucleotide polymorphism (SNP) associated with a variety of clinical conditions and susceptibilities [66], [67]. A group from Taiwan used SNP analysis to identify two SNPs in DAZL[15]. One SNP, located within the RRM, is associated with abnormal semen quality, while a second SNP, located outside of the RRM, was not associated with defects in spermatogenesis. Therefore, men with abnormal semen analyses may have other SNPs in their DAZL or DAZ genes. Men in central Texas have not been characterized for SNPs in relation to abnormal semen analysis. Another goal of this study was to perform SNP analysis in men with reported abnormal semen analysis in hopes to find a correlation between SNPs and reduced fertility in patients. Ultimately, haplotype analysis on a large cohort of patients will yield the most meaningful results.

Table 2. Estimation of Frequency of Y Chromosome DAZ Microdeletions in Relation to Sperm Cell Concentration of Semen. Publications cited utilized PCR primers found in the RRM of the DNA sequence for DAZ.

Ref	Date	Azoospermia†		Severe Oligozoospermia†		Oligozoospermia†		Fertile†		Total Patients	Locus
		Deletion	Total	Deletion	Total	Deletion	Total	Deletion	Total		
[12]	1995	21	89					0	90	179	sY254 directly and Includes sY254
[41]	1996			2	2					2	sY254 directly
[42]	1996	0	40	0	9			0	16	65	Includes sY254
[43]	1996	1	51	2	38	0	11	0	100	200	sY254 Directly & sY149
[44]	1996	0	19			0	14†	0	10	43	Includes sY254 by sY149
[8]	1996			13	370			0	200	570	Includes sY254
[40]	1997	3	74	2	94			0	86	254	sY254 Directly
[45]	1997	1	26	1	30	0	42	0	302	400	sY254 Directly
[46]	1997	1	13	0	79	0	79	0	33	204	sY254 Directly
[47]	1997	4	108	3	32	0	10	0	10	160	sY254 Directly
[48]	1997	0	19			7	111†	0	34	164	sY254 Directly
[49]	1997	6	25	2	61			0	35	121	sY254 by sY255
[50]	1997	5	43	0	28	0	87	0	55	213	sY254 Directly
[51]	1998	3	43	3	86	0	39	0	135	303	sY254 Directly
[52]	1998	8	50			2	136†	0	100	286	sY254 Directly
[53]	1999	1	22	2	68	0	27	0	27	144	sY254 Directly
[54]	1999	10	83	3	39	0	7	0	52	181	sY254 Directly
[39]	1999	10	36	26	278	5	200	4	920	1434	sY254 Directly
[55]	1999	14	55	14	85	0	40	0	100	280	sY254 Directly
[56]	1999	3	21			3	32†	0	35	88	sY254 Directly

Table 2. Continued.

Ref	Date	Azoospermia†		Severe Oligozoospermia†		Oligozoospermia†		Fertile†		Total Patients	Locus
		Deletion	Total	Deletion	Total	Deletion	Total	Deletion	Total		
[57]	1999	7	40					0	14	54	sY254 Directly
[58]	2000	1	15	0	50	0	62	0	141	268	sY254 Directly
[59]	2000	8	57	0	71					128	sY254 by sY149
[60]	2001	1	44	4	83			0	30	157	sY254 Directly
[61]	2001	5	54							54	sY254 by sY255
[62]	2001	2	5	5	14					19	sY254 Directly
[63]	2002			14	77			0	107	184	sY254 Directly
[64]	2003	8	142	1	38						sY254 Directly
[65]	2003	2	29	2	31			0	12		sY254 Directly
Tot.		125	1203	99	1663	17	897	4	2632		
%		10.39%	5.95 %	1.90%	0.15 %						

†Azoospermia indicates no sperm found in the semen, severe oligozoospermia indicates a sperm number concentration of up to 5 million sperm cells / ml, oligozoospermia indicates between 5 and 20 million sperm cells / ml, and normal is above 20 million sperm cells / ml.
‡These references were pooled because the authors did not identify patients specifically as severe oligozoospermic. Therefore the more conservative description of oligozoospermic was utilized.

Materials and Methods

PCR

Polymerase Chain Reaction (PCR) was performed using standard techniques where the PCR reactions contained 1X PCR Buffer (50mM KCl, 10mM Tris-HCl, pH 8.3, 0.001% gelatin), 2.5mM $MgCl_2$, 1.25mM each dNTPs, 0.5 Units/µl AmpliTaq Gold, and 5.3 µM of each primer. The reagents used for each of the PCR reactions were obtained from Applied Biosystems (Foster City, CA). PCR cycle parameters were specific to each product and determined empirically. The template used for the PCR reactions is described in the following section entitled "sY254 Screening."

sY254 Screening

For surveying the presence or absence of the sY254 locus, an anonymous submission procedure was used to obtain and evaluate cellular residues of semen samples from a population of male partners of infertile couples referred from a general population to a subspecialty practice of Reproductive Endocrinology and Infertility at Scott and White Hospital and Clinics (Temple, TX). This allowed for complete sampling of this patient population. The infertility practice is located in a central Texas multi-specialty clinic that provides healthcare services to a 16 county region of diverse racial, ethnic, social, and economic classes. The anonymous sampling procedure was established in 1996 and has added about 200 new patients each year. Males provide samples to the Scott & White Andrology Laboratory for semen analysis as a part of

evaluation for infertility. Men who were fertile or who had previously undergone vasectomies or surgical and medical interventions that reduced fertility, such as treatment for various cancers, were excluded. When the requested semen analysis was completed, the residual sample was cryopreserved for future study Only one sample per individual was saved. This residue was diluted in sterile saline and centrifuged at 800 x g for 20 min to produce a cell pellet. The supernatant was removed and a small quantity of sterile water was provided to wet the pellet to allow transfer to a sterile plastic cryovial. The vial was labeled with a unique random number from a computer-generated list that does not allow for repeats. The semen analysis results and certain demographic information such as patient age and smoking status were recorded on a form that included a random number as the only identifier. Samples were frozen and stored at -80°C. Results of the semen analysis were posted to a Microsoft Excel database along with results from molecular-based assays. At the time of this trial, the inventory of samples represented more than 1,300 individuals, 6 years of collections, and projected annual increases of 200 to 300 unique males each year.

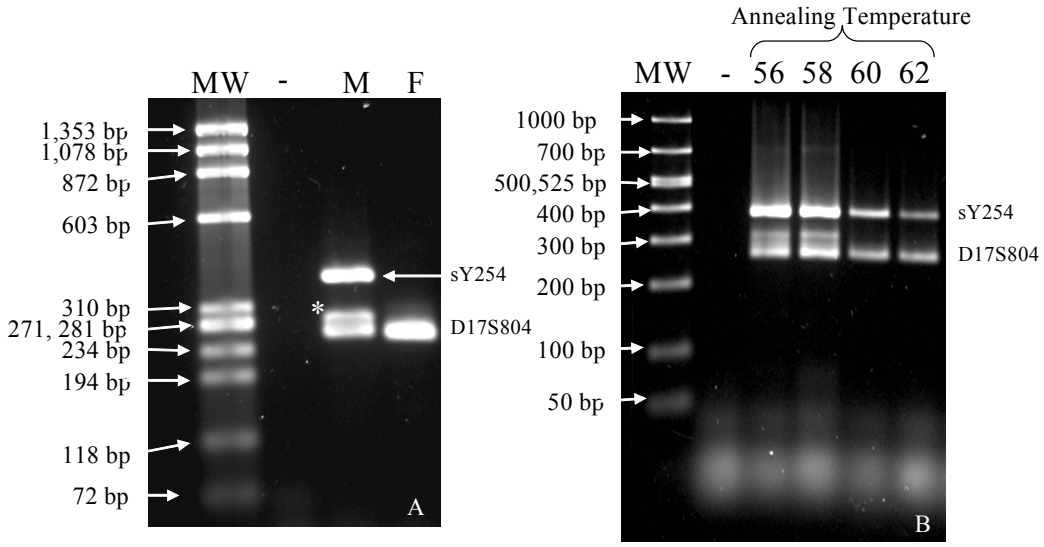


Figure 7. 2% Agarose Gel with SYBR®-Green Labeling of PCR Products in Screening Assay for sY254 Marker of Y Chromosome DAZ Locus. A. Lane 1: molecular size markers, Lane 2: negative control, Lane 3: male DNA (380 bp product of DAZ and 272 bp product from chromosome 17) with extra band on top, Lane 4: control female DNA (272 bp product of chromosome 17 marker). B. Gel showing that an increase in annealing temperature can get rid of the extra band found from the D17S804 primers. Lane 1: molecular weight standards, Lane 2: negative control, Lanes 3-6: an increase in annealing temperature gets rid of the extra band found in D17S804.

The DNA was extracted using the GenElute Kit (Sigma Chemical Co., St. Louis, MO) and assayed by PCR using primers for sY254 that map to the DAZ locus and primers for D17S804 that map to chromosome 17 as a control. DAZ primers for sY254 produce a 380 bp product with 5'-GGG TGT TAC CAG AAG GCA AA-3' and 5'-GAA CCG TAT CTA CCA AAG CAG C-3' while the D17S804 primers produce 272 bp product with 5'-TGT TCC ACG TGT AGG TGT CA-3' and 5'-CAC TGT GAT GAG ATG TCA TTC C-3'. The PCR products were separated by using an agarose submarine gel (Figure 7).

Additionally, in this series, 88 DNA extracts were tested using Y Chromosome Deletion Detection System (Promega Corp, Madison, WI), a commercially available multiplex PCR system for screening multiple markers of the Y-chromosome . Manufacturer's guidelines were followed. When specific PCR product bands were absent, individual PCR assays were performed to verify the results.

SNP Analysis

For SNP analysis of the DAZL RRM, DNA was extracted from 32 residual semen samples using the GeneElute kit. The samples were placed into 96-well plates and supplied to Polymorphic DNA Technologies (Alameda, CA) for processing. This company was used to minimize the cost of performing the sequence analysis on multiple patients. While the samples were in transit an e-mail was sent to Polymorphic DNA containing a text file with the DAZL genomic DNA sequence to be evaluated. Resulting sequences were examined using Microsoft Excel and DSGene from Accelrys (San Diego, CA). The approach to primer selection for the PCR reactions included the use of multiple primers sets to amplify the region of interest. Because of the similarity between DAZL and DAZ, only sequences which only contained DAZL sequence and no Y chromosome DAZ sequence were used in the final sequence analysis.

Results

DAZ Deletion Screening

A total of 1,325 semen residues were collected by the Andrology Laboratory at Scott & White, Temple, TX and results of semen analyses are summarized in Table 3. About 30% of samples had some type of abnormality based on sperm concentration.

Table 3. Semen Analysis Results of 1,325 Male Partners of Infertile Couples.

Characteristic	Units	Mean ± SD	Median	Range	Percent outside reference range *
Semen volume	**mL**	3.7 ± 2.4	3.1	0.1 – 18.6	24
Sperm concentration	**Millions/ mL**	51± 76	38	0 –2,200	29
Progressive motility	**%**	43 ± 20	43	0-96	21
Normal morphology	**%**	18± 9	17	0-61	43
Total number of sperm cells in semen sample	**Millions**	158 ± 161	112	0-1,452	22

*Samples outside of reference range for volume (<2.0 mL), sperm cell concentration (<20 million/mL), progressive motility (<25%), normal morphology using strict criteria (<15%), total sperm cells in ejaculate (<40 million)[3].

Only 450/1325 (34%) samples were within reference range for all characteristics.

According to WHO standards[3], the most common abnormality is reduced concentrations of sperm cells in the ejaculate related to, but not identical to a reduction in the total number of sperm cells in the sample. Of the semen residues available for analysis, 669 were extracted and subjected to PCR for analysis for DAZ deletion. Two of these samples were shown by PCR to be deficient in the sY254 locus in the RRM of DAZ. One sample contained 0.4 million and the other contained 17 million total sperm cells. Tables 4 and 5 summarize results of the DAZ deletion screening analyses based

on sperm cell concentration, the most commonly reported result, and total sperm cell content, a more stringent classification system for oligozoospermia. None of the 43 semen samples lacking sperm cells were deleted for the sY254 locus. Although information from clinical examinations was not available to verify that these men had unobstructed azoospermia, none had a history of vasectomy, vasectomy reversal, or cancer treatment. However, the frequency of genetic abnormalities associated with congenital absence of the vas deferens was not assessed, so some men might have absence of the vas deferens and, therefore, suffer from obstructed azoospermia. In this case, semen volume would be expected to be in the range of 0.5 ml. However, all men tested in the current study had semen volumes of 1.0 ml or greater. With this limitation, these samples may not include men with unobstructed azoospermia. Nevertheless, the frequency of 0/43 might underestimate the actual population frequency.

The frequency of DAZ deletions in male partners of infertile couples in this sample of central Texas patients (Table 4) is significantly less than published results (Table 2). In men with normal concentrations of sperm cells in their semen, the frequency of microdeletions in this series was low and similar to that reported in the literature.

Table 4. Relationship of Y Chromosome DAZ Microdeletion to Sperm Cell Concentration Results for Male Partners of Infertile Couples in Central Texas. All samples were screened for the presence or absence of the sY254 locus.

Sperm concentration (million/mL)	*N*	Frequency in Population (%)	DAZ deleted / Screened (%)	95% CI	Reported estimates
"0" (azoospermia)	51	3.8	0/43 (0%)	0 to 8.2%	125/1203 (10.4%)
> 0 to < 5	66	5.0	2/63 (3.2%)[†]	0.6 to 12.0%	99/1663 (6.0%)
5 to < 20	264	19.9	0/201 (0%)	0 to 2.3%	5/604 (0.8%)
20 and greater	944	71.3	0/362 (0%)	0 to 1.0%	4/2632 (0.2%)
Total	1325	100	2/669		

[†]Sample 729 with 4 million/mL and sample 237 with 0.089 million/mL

As DAZ deletions from the Y chromosome relate most closely to sperm cell production, comparing the frequency of deletions in association with sperm cell concentration can be misleading. For example, some men may have low numbers of sperm in their ejaculate, but a reduced volume of seminal vesicle secretions results in a normal concentration. Others may have normal numbers of sperm with high volume of semen resulting in reduced concentrations. Therefore, results from this series based on sperm cell content in the total ejaculate are summarized in Table 5. Using this system the two males with DAZ deletions can be classified as having either severe oligozoospermia or oligozoospermia (Table 5). Other reports in the literature have not been based on this system and have not provided semen volumes to allow similar

comparisons. Table 6 summarizes results from multiple tagged sequence testing of #237 male deleted for DAZ using a commercially-available kit. Semen residue #729 did not provide enough DNA after repeated analyses to support additional multiple sequence testing. However, sample #237 was used for this analysis. The tagged sites, sY152 through sY157, included five DAZ-specific sites found to be deleted. Other regions of the Y chromosome associated with the AZFa and AZFb portions were present in genomic DNA from this male. This supports the observation that the male providing the sample #237 has a palindromic deletion (P1.1 to P3) as described by Repping et al. [36]. Figure 8 shows the regions amplified with PCR using the Promega kit.

Table 5. Relationship of DAZ Deletion to Total Number of Sperm Cells in Semen Samples of Male Partners of Infertile Couples in Central Texas.

Total sperm cells (millions)	***N (% Population)***	**% DAZ deleted**	**95% CI**
"0" azoospermia	51 (3.8%)	0/43 (0%)	0 to 8.1%
> 0 to < 5	31 (2.3%)	1[†]/31 (3.2%)	0.8 to 21.4 %
5 to <20	74 (5.6%)	1[††]/71 (1.4%)	0 to 5.1%
20 to <40	135 (10.2%)	0/103 (0%)	0 to 3.6 %
40 and greater	1034 (78.1%)	0/421 (0%)	0 to 0.9 %

[†] Sample 237 with 398,000 sperm cells
[††] Sample 729 with 16.8 million sperm cells

Table 6. PCR Detection of Y Chromosome Deletions*.

Primer set	Y chromosome site	Presence of PCR product from semen residue #237
SY81	DYS 271	+
SY181	KAL-Y	+
SY121	DYS212	+
SYPR3	SMCY	+
SY124	DYS215	+
SY127	DYS218	+
SY128	DYS219	+
SY130	DYS221	+
SY133	DYS223	+
SY145	DYF51S1	+
SY153	DYS237	+
SY152	DYS236	-
SY242	DAZ	-
SY239	DAZ	-
SY208	DAZ	-
SY254	DAZ	-
SY255	DAZ	-
SY157	DYS240	-

* Y chromosome microdeletion screening with Y Chromosome Deletion Detection System (Promega Corp, Madison, WI), a multiplex PCR system for screening multiple markers of the Y-chromosome.

Illustrated are the three AZF regions, the five major palindrome locations, and the various STS sites used in the multiplex Promega assay. The STSs highlight the repetitive nature of the Y chromosome in this region. The sequences used to generate this map are from Kuroda-Kawaguchi et al. and the sequence can be found in online supplemental material[33]. Figure 9 illustrates a magnified portion of the AZF region focusing mainly on AZFc and the STSs used primarily in PCR detection to identify the structure of deletions on the Y chromosome.

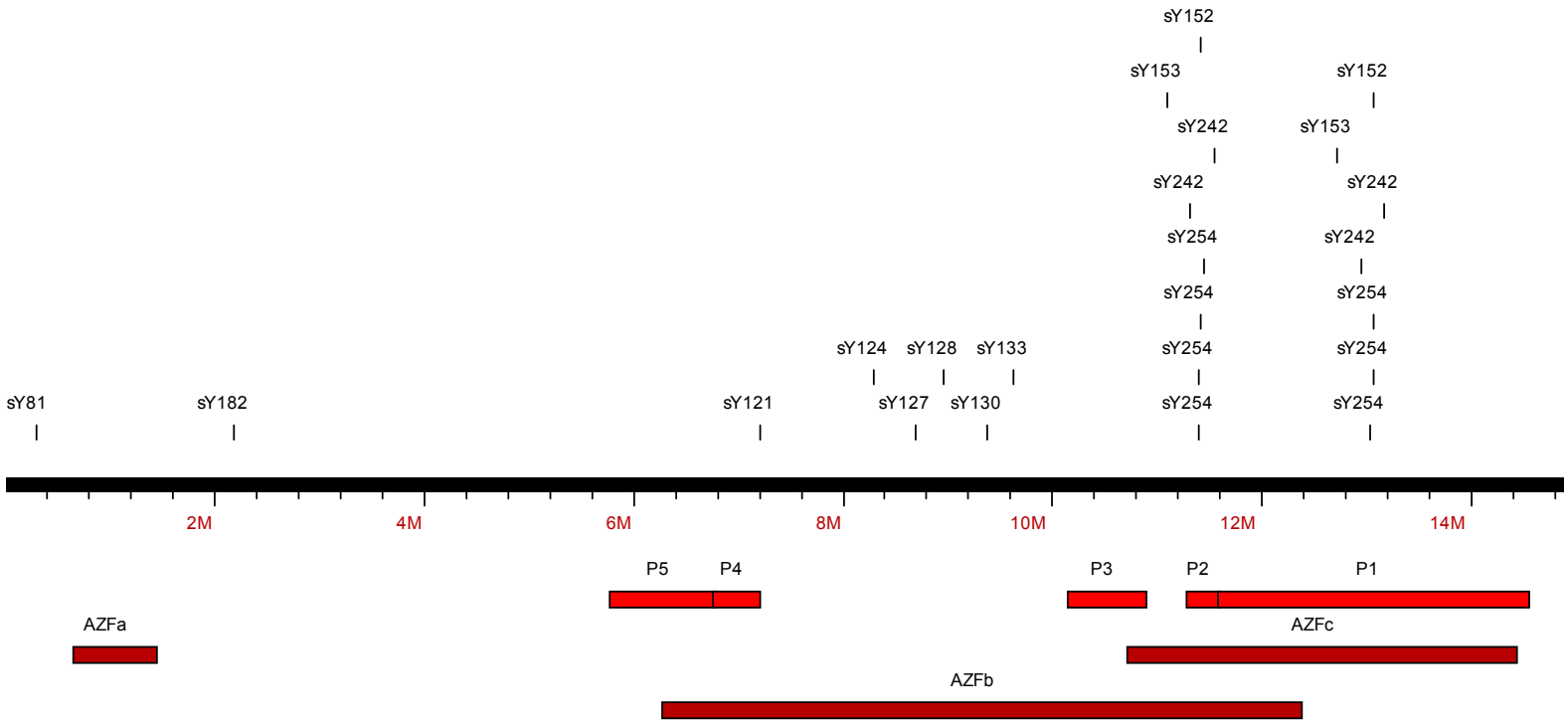


Figure 8. The AZF Regions and STS Sites on the Long Arm of the Y Chromosome. The three AZF regions are shown on the bottom as darker red bars where the palindromes are brighter red and shown just above. Above the black line showing distances are the locations of each of the STS sites found in the Promega kit.

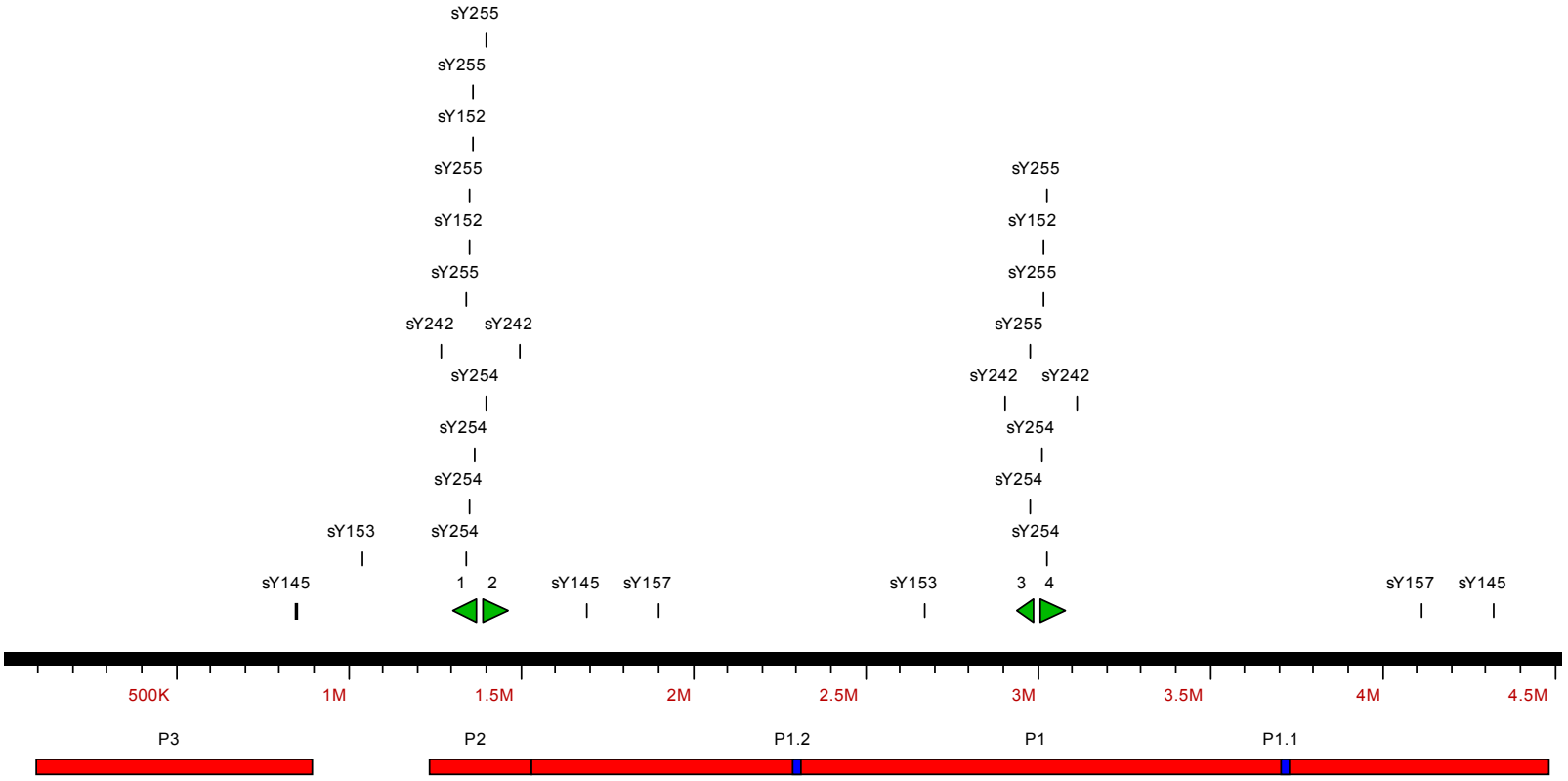


Figure 9. AZFc Region and STS Sites. The palindromes are located on the bottom of the figure with small blue boxes indicating minipalindromes which are hotspots for recombination; deleting portions of the DAZ copy number. Above the black line are the STSs found within the various AZF regions demonstrating the locations of multiple copies. The green arrows represent the DAZ copies found on the Y chromosome.

Are There Point Mutations in the DAZL RRM?

Another type of genomic mutation that could manifest itself in the form of aberrant function of DAZ proteins is the point mutation. The DNA from semen samples from 32 anonymous patients was subjected to sequence analysis of the RRM genomic DNA

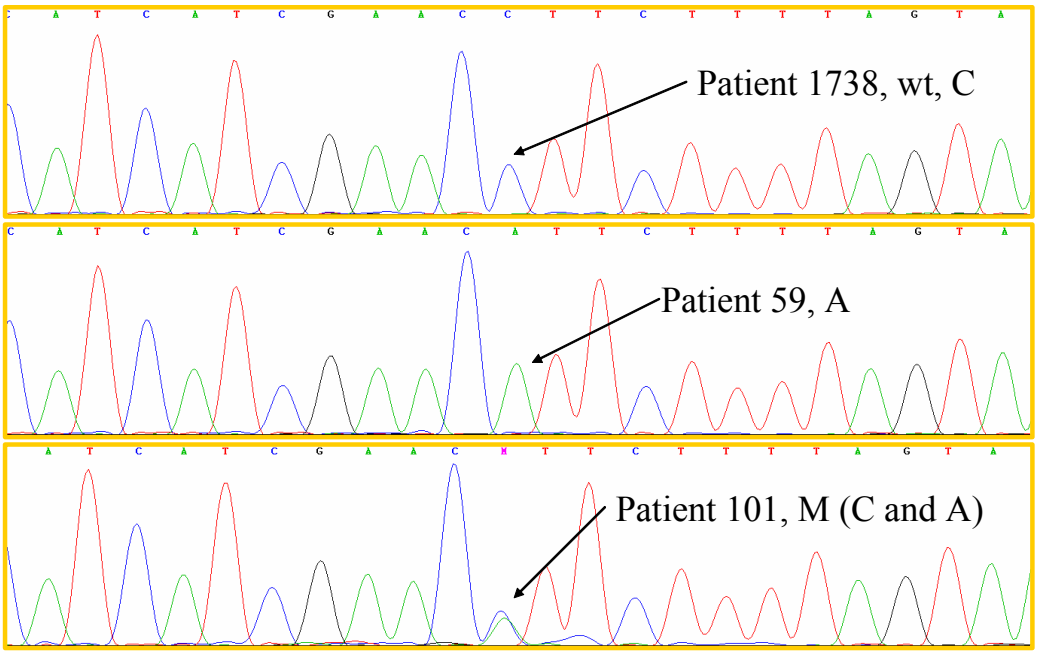


Figure 10. A Sample Comparison of Chromatographs from Different Patients. This figure represents SNP#2 found in Table 7. Patient 1738 represents "wild type" at this position, which contains cytosine. Patient 59 possesses an adenosine at the position, and Patient 101 contains both forms.

portion of the DAZL gene. The DAZL gene was chosen instead of the DAZ gene for two reasons. The first reason that DAZL was chosen instead of DAZ is that it was already stated that the DAZL is necessary for spermatogenesis. There will be no

spermatogenesis if there is no DAZL gene. In addition, detecting an SNP in one of the many copies of the DAZ gene would prove to be a difficult task. If there were an SNP in one of the DAZ genes, the result would be hidden from view by the other copies. The RRM was chosen because it is the defining feature of a group of RNA-binding proteins containing two RNP (ribonucleoprotein) sites. Because of the cost of sequencing and the difficulties in designing primers, only the RRM was focused on as a site for SNP exploration. The hypothesis of the experiment was that a point mutation leading to an amino acid substitution in the RRM within DAZL would change the function of DAZL due to modifications of structure and binding of DAZL to RNA.

The DNA sequence chromatograms were examined for the presence of SNPs within the area of exons in the RRM of DAZL. The size of the genomic fragment analyzed was 1,508 bp in genomic DNA on chromosome 3. Figure 10 provides an example of patient sequences aligned with wild-type through a region of interest within the RRM. It illustrates what results look like when sequences are aligned with each other and for each of the possible nucleotides at each position. The SNP was found at position 262 of the genomic DNA complement of the RRM, which localizes to an intron. Patient 1738 has a cytosine while patient 59 has an adenosine at this position, and patient 101 had equal amounts of cytosine and adenosine at a single position. The double signal at position 262 of the RRM for patient 101 indicates equal amounts of template for detection in the PCR reaction. This suggests that the sperm cell population contains equal numbers of sperm with each SNP at this position. Because there were 14 other patients exhibiting this genotype it is unlikely that this is an artifact of the PCR reaction.

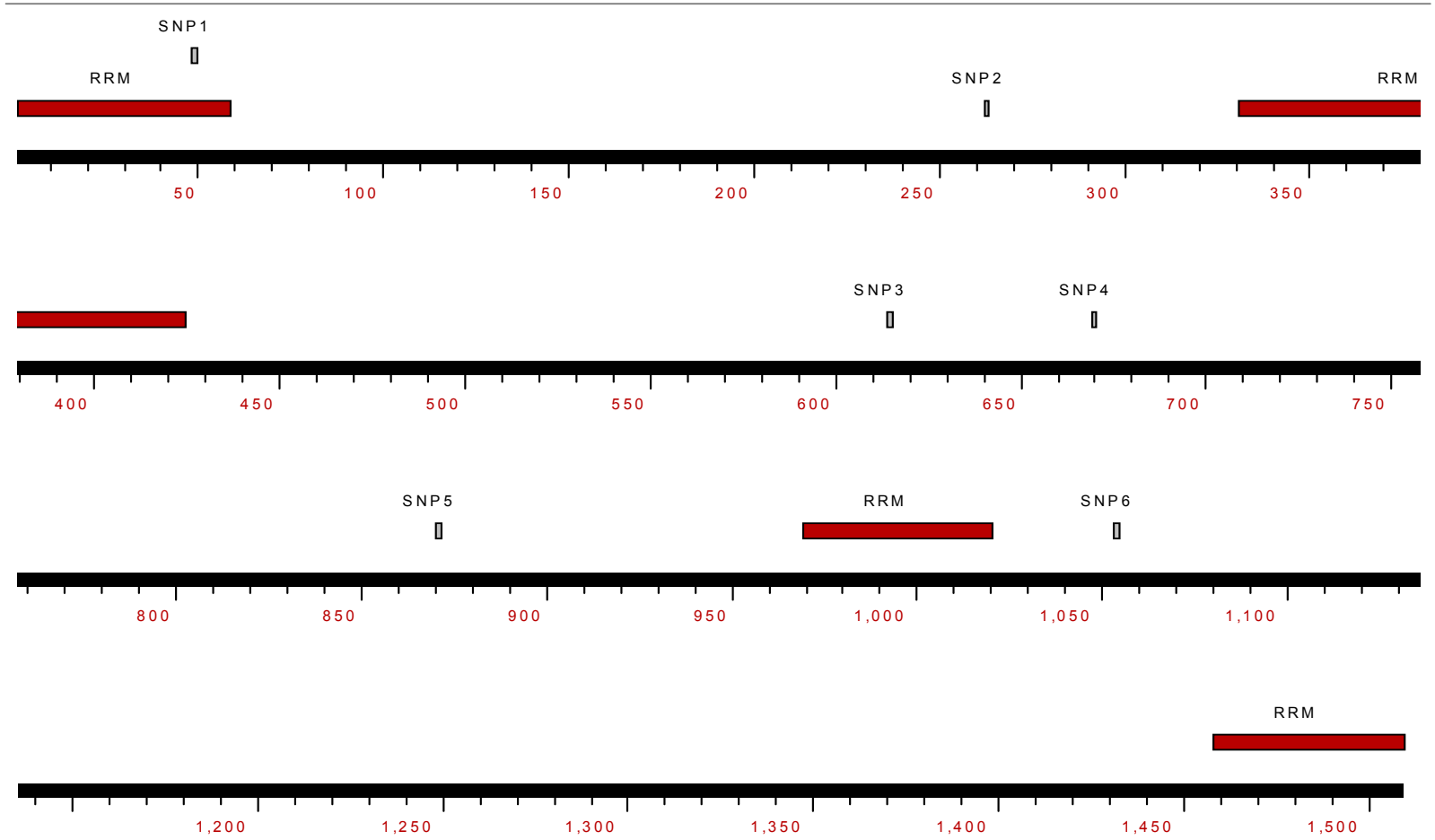


Figure 11. DAZL RRM Genomic SNPs. The location of the various SNPs within the genomic DNA comprising the RRM of DAZL are indicated along with the exons depicted as red horizontal bars. The size of the compete genomic complement of the RRM is also shown as 1508 bp. All SNPs which were discovered were found within introns except SNP1 which was a silent or null mutation.

Table 7. Summary of Distribution of SNP Types.

SNP	Intron/Exon	Nucleotide(s)	# Patients	Mutation/Transversion	# Normal	# Oligozoospermic	# Azoospermic
SNP#1	Exon	T (wt)	17		3	11	3
		W (A and T)	1	ATT (Ile) - ATA (Ile)	0	0	1
SNP#2	Intron	C (wt)	3		1	2	0
		M (C and A)	14	None	0	10	4
		A	11	None	1	8	2
SNP#3	Intron	A (wt)	23		3	15	5
		C	1	None	0	1	0
SNP#4	Intron	G (wt)	21		1	15	5
		K (T and G)	1	None	0	1	0
SNP#5	Intron	A (wt)	21		3	14	4
		R (A and G)	10	None	0	8	2
SNP#6	Intron	A (wt)	11		1	6	4
		M (C and A)	5	None	0	4	1

There were a total of six SNPs detected in the genomic region of 32 patients within the RRM of DAZL. Of the six SNPs detected, only SNP#1 was located within an exon while the remainder were within introns. Within SNP#1 the only mutation found was a silent mutation- not leading to an amino acid change. The remainder of the SNPs found was within introns and found not to affect splicing of the DAZL protein. Sperm numbers are also indicated where normal is considered >20 million sperm cells / ml, oligozoospermic is considered up to 5 million sperm cells / ml, and azoospermic is considered no sperm cells in the semen.

The resulting SNPs were arranged graphically (Figure 11) to reveal the absence of SNPs within any exon, except for one SNP in an exon resulting in a silent mutation that presents as a polymorphic nucleotide without a resulting change in amino acid sequence. In this case there was an isoleucine to isoleucine conservative change (ATT to ATA). The location of SNPs within introns were unlikely to have a direct effect on splicing signals present in the genomic complement of DNA which could potentially disrupt the RNA binding domain of DAZL. Four sequences essential for splicing include the 5' splice site, the 3' splice site, the branch site, and the polypyrimidine tract [68]. The 5' splice site possesses the sequence Exon+AG|**GU**RAGU where the splicing

site is delineated by the vertical bar. The two bold nucleotides which follow the bar are highly conserved and rarely, if ever, change. The key to the nucleotides is: R represents purine nucleotides, Y represents pyrimidine nucleotides, and N is any nucleotide. The 3' splice site contains the sequence Y**AG**|+Exon where the vertical bar precedes the following exon. The branch site, found between 18 and 40 nucleotides upstream of the 3' splice site, contains the sequence YNCUR**A**Y where the branch formation site is underlined. Finally, the polypyrimidine tract Y_n precedes the 3' splice site in a distance that is variable. By examining sequence data directly, no SNP in the DAZL region contained a mutation in any of these sites. Furthermore, when the genomic sequence, imbued with SNPs garnered from this experiment, was applied to the exon prediction program Genscan (http://genes.mit.edu/GENSCAN.html) it was found that no SNP altered the predicted peptide sequence. To organize the data, Table 7 lists the number of individuals with each type of SNP. All individuals had at least one type of SNP. Additionally, semen characteristics are indicated for each of the SNP groups.

Extensive SNP analysis on patients within the DAZL region created haplotypes from which susceptibility to spermatogenic failure can be associated[69]. When their data were compared to that from the present study, SNP6 was the same nucleotide found to be polymorphic within the population studied by Teng (2006). The nucleotide in question was not found, however, to be associated with any type of spermatogenic failure.

Discussion

While numerous publications have documented the frequency of microdeletions in the DAZ region of the Y chromosome in males referred to centers for specialized assisted reproduction procedures, few reports indicate results of assessment of the frequency in a general population of male partners of infertile couples. Kent-First et al. 1999, evaluated 514 infertile males with 7% of 200 of the men from an unselected infertile population described as a consecutive series referred to a urology practice for male factor infertility and having microdeletions[39]. Subgroups of highly selected patients in this study had between 20.5% and 58.3% of individuals with microdeletions. These results indicate that highly selected subgroups of men have higher frequencies of microdeletions than the general population of infertile subjects. Male partners of infertile couples in central Texas included 1300 individuals from which the present study provides results from screening 669 samples.

Earlier reports suggested frequencies of DAZ deletions of 11% among males with no sperm in the ejaculate and 6% for males with sperm concentrations up to 5 million/ml (Table 2). These results were used to support the recommendation that laboratories offering semen analysis for evaluation of male partners of subfertile couples include PCR-based screening methods. However, in the present study, over a period of 6 years, only 1300 individual males had an initial semen analysis. Of these, only 2 of 669 samples were screened for DAZ deletion using optimized PCR with STS marker sY254 to detect a deletion. Both semen samples contained sperm cells and one sample

had sufficient numbers of motile sperm to suggest a good likelihood of success with assisted reproduction procedures. Because of the anonymous nature of the sampling process, a follow-up assessment was not possible. However, the frequency of affected males is rare in this study's population. Singh et al. (2005) stated that their patients had no AZFa deletion whereas reports from others in India indicated AZFa deletions correlated with reduced fertility in males, perhaps due to, "…different backgrounds and effect modifiers. Since male infertility is a multifactorial disorder, the contributions of environmental and occupational insults may not be underestimated."

The sY254 marker is utilized to assess the presence of genomic copies of DAZ because of its location in the expressed portion of the mRNA binding domain. At least eight such domains exist in the four copies of DAZ located on the Y-chromosome, so the absence of a PCR signal suggests that all copies of Y-chromosome DAZ have been lost. However, the presence of the PCR signal does not prove that the DAZ genes will result in expression of functional protein which adds complexity to the interpretation of screening data for this single Y-chromosome gene locus.

In addition to relatively large changes in the final structure of the DAZ proteins, subtle genomic mutations may have a measurable effect on fertility in humans. One such example is a point mutation where a single nucleotide causes a single amino acid substitution in an area critical for protein function. Diseases commonly associated with SNPs are cystic fibrosis, Creutzfeldt–Jakob, and Crohn's disease [67], as well as Alzheimer's Disease [70] in which SNPs in brain-derived neurotrophic factor influence development of the disease. As mentioned in Chapter I, a mutation in the androgen

receptor can also have a dire consequence on male fertility. Furthermore, a SNP in the protamine 1 (PRM-1) gene is a candidate for infertility in the male [71]. The SNP within the PRM-1 gene disrupts an arginine-rich conserved core of the protamine needed for binding DNA. Additionally this SNP may add another phosphorylation site within the protein to affect secondary structure of the protein.

In the current study, a mutation in exon 1 of the RRM of DAZL was identified and proved to be a silent mutation encoding isoleucine in both cases. The goal of SNP screening was to find a point mutation associated with reduced fertility due perhaps to conformational changes in the encoded protein. The other SNPs in genomic DNA of the 32 patients were confined to introns. One mechanism by which SNPs in introns can affect fertility is alteration of splice sites that affect how the DAZL protein is assembled. An example of the effect of a SNP upon splicing is the alternative splicing observed from a SNP in humans with reduced nicotinamide adenine dinucleotide phosphate:quinone oxidoreductase-1 variant 2. When this SNP is present, exon 4 is removed from the gene leading to expression of a truncated, nonfunctional gene product[72]. In the present study, SNPs were located in regions remote from known splice junctions or signals.

Recently, Teng et al. performed a more comprehensive SNP analysis within the DAZL region on the Han population in Tiawan[69]. They found a SNP within an intron of DAZL dense with polypyrimidines called the PCE or polypyrimidine controlling element. This was interesting because Zuccato et al. reported that a mutant within the PCE determined whether exon 9 was present within the cystic fibrosis transmembrane

conductance regulator gene[73]. Because Teng et al. found SNPs outside of the RRM, it is important for to examine sequences outside of the RRM and cover the complete genomic sequence for mutations that may lead to changes in functional proteins. Because of potential splice variants, RT-PCR should be used to identify potential exons that could be removed from the total DAZL protein.

More patient samples are needed to identify rare SNPs that may affect the DAZL gene. When more data are present, haplotypes can be formed and linked to specific semen analyses to establish association with sperm production. However, there may be more levels of regulation when Y chromosome DAZ's function is included in spermatogenesis as DAZ can exist in several forms that have greater or fewer repeat regions important for protein-protein interactions. A number of proteins that interact with DAZ directly have been identified [17, 74, 75]. The DAZAP1 codes for an RNA-binding protein, whereas DAZAP2 has no known function. Both of these proteins bind DAZ proteins through the repeat region and are expressed in the testis. The DAZAP1 is also expressed in ovarian luteal cells and may have functions similar to DAZ proteins in that they bind RNAs related to cell cycle control. The DZIP1 also binds DAZ proteins and is expressed in premeiotic spermatogonial cells. Because of the potential for interactions of DAZ with other proteins, even single amino acid changes and small changes in secondary structure may affect binding to RNA or other proteins to affect downstream functions of DAZ. To assess function, the proteins encoded by the various forms of DAZ must be expressed and experiments performed to determine specific functionality.

CHAPTER III

ISOLATION, CLONING, AND VERIFICATION OF DAZL PROTEINS AND VARIANTS

Introduction

Genomic DNA can be used to infer amino acid primary sequence of a protein, but not function. It is necessary to express the DAZL and DAZ genes in the form of proteins to assess their function. It has been suspected for some time that the DAZ family of proteins are RNA-binding proteins[7, 76]. DAZ proteins have been expressed in *E. coli* systems and used in yeast for two hybrid assays[38, 77-81] and a variety of other assays have been utilized to determine their function. These include the electrophoretic gel shift mobility assay (EMSA) which is commonly used for determining equilibrium and kinetic constants of RNA-protein interactions[82]. Other assays to assess RNA-protein interactions include RNA footprinting, nitrocellulose filter binding, and northwestern screening of libraries[83]. In this study DAZL, mutant forms of DAZL, DAZ, and MBP proteins were cloned, expressed, and tested for binding to radiolabeled synthetic RNA to assess RNA binding activity.

Materials and Methods

PCR from Phage Library to Amplify DAZL

The basic construction of DAZL and all other proteins involved in this study made use of the TEV protease linker on the amino terminus of the protein. A TEV protease linker was incorporated into the protein such that upon cleavage of the desired protein, there would be only an additional glycine on the amino terminus. Glycine was chosen to minimize potential effects on protein function due to its small size and neutral charge. Shown below in Figure 12 is an example of the 5' end configuration when designing primers were used to amplify a gene with the TEV protease linker. The premise is that the 3' end of the primer will anneal to the sequence of interest and the amplifications during PCR will amplify not only the gene of interest but will include, on its 5' end, the signal to create a TEV protease site.

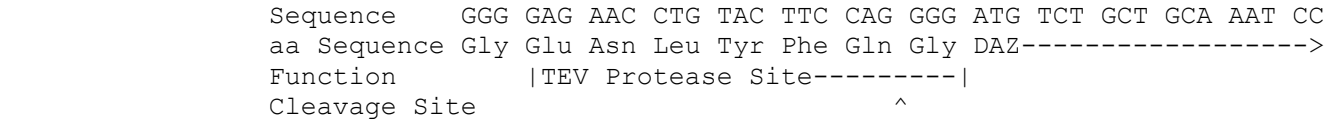


Figure 12. Illustration of TEV Protease Linker.

PCR conditions used to generate a DNA copy of the DAZL coding sequence from the Clontech phage library included a final concentration of 2mM $MgCl_2$, 200µM

each dNTPs, 530µM each primer, 1X PCR buffer (see Chapter II Materials and Methods), and 1.25 units of AmpliTaq Gold. All reagents were obtained from Applied Biosystems (Foster City, CA). The primers (designed specifically for this procedure) were DAZLF+TEV- 5'-GGG GAG AAC CTG TAC TTC CAG GGG ATG TCT ACT GCA AAT CCT GAA-3' and the reverse primer DAZLR+3Stops- 5'-TCA GCT AAT TAT CAA ACA GAT TTA AGC ATT GC-3'. The PCR was performed in Stratagene's Robocycler (La Jolla, CA) with an initial denaturation of 95°C for 15 min then 35 cycles at 94°C for 30 sec, 56°C annealing for 1 min, and a 72°C extension for 1 min and 30 sec. A final extension for 5 min at 72°C finished the PCR. 5 µl of 6X loading dye consisting of 2.5 g/ml Ficoll and 25 mg/ml bromophenol blue was used in a 25 µl PCR reaction. The agarose gel used for product detection was a 1% SeaKem LE (BioWhittaker Molecular Applications, Rockland, ME) gel run under standard conditions of 70V for 90 to 120 min in 1X TBE.

Cloning of PCR Products

Cloning employed the quick cloning protocol as described by Invitrogen using the pCR2.1 cloning vector and the following transformation into TOP10 cells (Invitrogen, Carlsbad, CA). Briefly, 4 µl of the PCR reaction was mixed with 1 µl 10 ng/µl of the pCR2.1 vector and incubated at room temperature for 5 min. Four microliters of the cloning reaction was mixed with the contents of a tube of TOP10 cells and gently mixed by flicking the tube softly. The transformation was allowed to proceed on ice for 30 minutes and 60 µl of the transformed cells were plated onto prewarmed and

dried LB/agar plates containing 50 µg/ml of carbenicillin. The plates were incubated overnight at 37°C.

*Streak Plate*s

Streak plates were made by picking single colonies from the cloning plate and streaking separate lines on LB/agar plates containing antibiotics. PCR was performed on individual isolates to confirm identity and the primers used were specific for each application.

Midipreps

Midipreps were performed according to manufacturer's instructions (Qiagen, Valencia, CA). The only modifications were to midipreps involving low-copy plasmids such as pMAL (New England Biolabs, Ipswich, MA). In these cases, the amounts of cells and denaturants were increased two-fold. However, all of the lysed samples were pooled for application to a single column for increase in DNA amount. In the majority of cases, 25 ml of an overnight culture was centrifuged for 15 min. in a Beckman JA-20 rotor at 4°C at 7,000 rpm in polycarbonate tubes. After centrifugation, the supernatants were discarded and the pellets were placed at -20°C overnight. The pellet was suspended in 4ml of QIAGEN Buffer P1. The pellets were homogenized with pipetting and vortexing until no visible clumps remained. Four milliliters of QIAGEN Buffer P2 was added to lyse the cells and mixed with gentle rocking by hand. The solution was then allowed to incubate at room temperature for no more than 5 minutes. If lysis was

not complete after 1 minute then more of buffer P2 was added until lysis was complete. Complete lysis of cells had the appearance of a translucent rather than opaque liquid in the tube. After lysis was complete, an equal volume of QIAGEN Buffer P3 to QIAGEN Buffer P2 was added to the tube and rocked gently by hand. The solution containing QIAGEN Buffer P3 was incubated on ice for 15 min. The samples were then centrifuged in a JA-20 rotor for 30 minutes at 4°C at 13,000 rpm. The supernatant was promptly poured off and re-centrifuged under the same conditions but only for 15 min. The subsequent supernatant was poured into a new tube.

During centrifugation, a QIAGEN-tip 100 was equilibrated with 4 ml of QIAGEN Buffer QBT, and the supernatant was applied to the column to let the DNA bind. If the midiprep made use of multiple tubes for lysis of the same construct, all of the tubes containing DNA were applied to the column. The column was washed two times with 10 ml of QIAGEN Buffer QC. The DNA was eluted with 5 ml of QIAGEN Buffer QF. For precipitation, 3.5ml of 100% -20°C isopropanol was added to the DNA-containing solution. The samples were centrifuged in a JA-20 rotor for 30 min. at 4°C at 12,000 rpm to pellet the precipitated plasmid DNA. The supernatants were carefully poured off and the pellet was washed with 2ml of room-temperature 70% ethanol and immediately centrifuged again for 10 min. The supernatant was poured off, and the pellet was air-dried for 30 min. then solubilized in 300 µl of QIAGEN Buffer EB.

Digests of pMAL and DAZL

Digests were performed on plasmids containing DAZL and the pMAL vector using EcoRI (New England Biolabs) so that the EcoRI fragment containing the DAZL cDNA could be inserted into the EcoRI site of the pMAL vector. With the exception of a single EcoRI site in the 3' UTR, the sequence of DAZL is not known to have any EcoRI sites. The reaction conditions for the DAZL plasmid were 7 µl of water, 2 µl of 10X NEB EcoRI Buffer (final reaction concentration was 50 mM KCl, 100 mM Tris-HCl, ph 7.5, 10 mM $MgCl_2$, and 0.025% Triton X-100), 2 µl of NEB EcoRI enzyme (Cat# R0101S, 20 Units/µl), and 9 µl of the 312 ng/µl DAZL template DNA in a solution of 10 mM TE. The pMAL reaction contained 12 µl of water, 2 µl of the 10X EcoRI buffer, 1 µl of the EcoRI enzyme, and 0.5 µg of the pMAL C_2X plasmid which was supplied in a tube with 10 µg of DNA. The digests were incubated at 37°C for 3.75 hr. The pMAL digest was treated with CIAP (Calf Intestinal Phosphatase) where 1 µl of 1 U/µl of diluted CIAP (5 µl of 25 U/µl CIAP + 120 µl dilution buffer) was added to the pMAL digest. The reaction was incubated at 37°C for 5 min. at which point 5 µl of 50 mM EDTA, pH 8.0 was added. The reaction containing the EDTA was incubated at 65°C for 15 min. After enzymatic digest, the product was electrophoresed on a 1% SeaKem GTG gel (BioWhittaker) and the bands of interest extracted with a scalpel. After excision from the gel, the DNA was recovered using the Genelute Minus Kit (Sigma) according to manufacturer's instructions.

Ligation Reactions

Ligation of the pMAL digest and the DAZL digest was performed using NEB's T4 DNA Ligase (Cat# M0202S, 400 U/µl). The reaction consisted of 2 µl of 10X T4 Ligase Buffer (1X Final Concentration of 50mM Tris-HCl, pH 7.5, 10mM $MgCl_2$, 10mM DTT, 1mM ATP, and 25 µg/ml BSA), 400U of T4 DNA Ligase, 16 µl of the digested insert, and 1 µl of the digested vector from above. The volumes of vector and insert were arbitrarily selected. The reaction proceeded at 16°C overnight.

Test Expression of Clones

Clones were tested for expression of protein with separate 5 ml cultures grown from picks from streak plates to an OD_{600} of 0.5 and induced with 0.3mM final isopropyl β-D-1-thiogalactopyranoside (IPTG), a molecular mimic of allolactose. Separate samples of 1 ml each were collected and the cells pelleted. Each pellet was dissolved with 2X SDS Buffer in Tris, pH 6.8 (30% glycerol, 10% SDS, 0.6M DTT, 0.012% bromophenol blue). The denatured proteins were electrophoresed on a 10% Tris-HCl BioRad Criterion gel (Hercules, CA). Protein gels were stained with Coomassie using standard molecular biological laboratory practices.

Freezer Stocks of Clones

Clones were grown to mid-log phase started from overnight cultures in LB with appropriate antibiotics at the proper concentrations, diluted to 50% with glycerol, and stored at -80°C.

Electroporation

Electroporation was performed using an Eppendorf Electroporator 2510 (Madison, WI) and BioRad Gene Pulser®/E. coli Pulser™ Cuvettes (Hercules, CA) set at 1900V for all samples. The procedure involved mixing DNA and electrocompetent cells together in an electroporation cuvette and applying the voltage. A high time constant for each pulsing run indicated a desirable run (~ 5.2 milliseconds).

Assessing Purity of DNA Samples

A 1:200 dilution of the sample to be read was made in 1 ml ddH_2O and absorbance was read at 260nm and 280nm. From these numbers the concentration and purity of DNA was calculated as concentration of DNA = A_{260} X dilution factor X 50 ng/µl. Purity of the sample was easily calculated by A_{260} / A_{280} with the expectation of a ratio in the range of 1.7 as ratio values of 2.0 or greater indicate ribonucleic acid contamination and ratios of 1.6 or less indicate protein contamination.

Sequencing Verification

All sequencing was performed by Lone Star Labs (Houston, TX) and compared with data from Genbank. After plasmid preparations were performed and the DNA evaluated for amount and purity, a sample was sent via overnight courier. If needed, primers specific for amplification were also sent conforming to Lone Star Labs' requirements for Tm, length, and concentration.

Site-Directed Mutagenesis

The mutagenesis procedure used to fix a mutation in the generated clone in DAZL was performed using Stratagene's QuikChange® Site-Directed Mutagenesis Kit (Cat# 200519-5, La Jolla, CA) according to manufacturer's instructions. Briefly, this method requires creating a non-methylated copy of the plasmid with a point mutation included in complementary primers. The chosen primers were based on size, specificity, and thermodynamic characteristics. Stratagene recommends that the primers be between 25 and 45 bases in length, the Tm's should be greater than 78°C, the desired mutation should be in the middle of the primer with at least 10 to 15 bases of correct sequence on each side, should have a minimum GC content of 40%, should terminate in one or more C or G bases, and should be purified by PAGE or HPLC. For the current study, 125 ng/5 µl of each PAGE-purified primer was added to the PCR reaction along with 5 ng/5 µl total input plasmid, 29 µl of water, 5 µl of 10X Buffer (final reaction concentration of 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 0.1% Triton X-100, and 0.1 mg/ml BSA), 1 µl dNTP mix supplied in the kit, and 1 µl of 2.5 U/µl Pfu Turbo. The PCR reaction was carried out in Stratagene's Robocycler with the following program: Step 1: 95°C for 30 seconds and 55°C for 1 minute, Step 2: 12 cycles of 95°C for 30 seconds, Step 3: 68°C for 8 minutes. After the PCR reaction, the methylated non-mutated DNA was degraded by the addition of 10 units of enzyme DpnI and the resultant nicked circle DNA was transformed into supercompetent XL-1 Blue cells. Primers specifically developed to repair mutations were: forward primer DAZL0.1 to 0.2F (5' CCT GTT GGG GAG CAA AGG AGC TAT GTT GTA CCT CCG GC 3') and

reverse primer, DAZL0.1 to 0.2R (5'-GCC GGA GGT ACA ACA TAG CTC CTT TGC TCC CCA ACA GG-3'.

In addition, published point mutations in DAZL were produced using this method. The primers were specifically developed from sequence data provided by Teng, et al.[15] who found two major mutations in the general Taiwanese population. Both SNPs encode an A to G transition encoding for a Thr to Ala amino acid alteration. The second mutation (labeled DAZLmut2) was in the RNA recognition motif, but the first mutation (labeled DAZLmut1) was not and preceded the RRM region in the sequence. The DAZLmut1 forward sequence was 5'- GCA AAT CCT GAA ACT CCA AAC TCA GCC ATC TCC AGA GAG G -3' And the reverse primer sequence was 5'- CCT CTC TGG AGA TGG CTG AGT TTG GAG TTT CAG GAT TTG -3'. DAZLmut2 forward primer sequence was 5'- GGA GGA ATT GAT GTT AGG ATG GAT GAA GCT GAG ATT AGA AGC -3' and the reverse primer was 5'- GCT TCT AAT CTC AGC TTC ATC CAT CCT AAC ATC AAT TCC TCC -3'. The methodology used in the PCR reaction and all associated calculations used in the previous example were used in creating these primers and PCR reaction.

Expression and Purification of DAZL Proteins

Stabs from freezer stocks were used to start 10 ml overnight incubations of each clone in LB containing 100 µg/µl carbenicillin at 37°C with shaking. The overnight solution of LB was seeded with 10 ml into a 2 L aeration flask containing 1 L of LB with 100 µg/µl carbenicillin and 2 g/l glucose to inhibit amylase formation which would

therefore inhibit the MBP fusion proteins from binding to the amylose resin. The culture was grown at 28°C with shaking to an OD_{600} of 0.5 which required about 5 hours. IPTG was added to a final concentration of 0.3 mM and expression of protein induced for 3 hrs. The cells were pelleted and stored at -20°C overnight. The pellets were suspended in column buffer (20mM Tris-HCl, pH 7.4, 200mM NaCl, and 1mM EDTA) and poured into 50 ml conical tubes, dounced 10X to reduced clumping of cells and solutions passed through a French Press two times. The lysate was clarified through centrifugation and the desired proteins were enriched using amylose resin with multiple washes with buffers according to manufacturer's instructions (New England Biolabs, Ipswich, MA).

Cloning of DAZ+TEV from Clontech Phage Library

Specifically developed primers were synthesized by IDT (Coralville, IA). The forward primer sequence, 'YDAZ+TEV' to clone Y chromosome DAZ was 5'- GGG GAG AAC CTG TAC TTC CAG GGG ATG TCT GCT GCA AAT CC- 3'. The reverse primer, YDAZ2&3R, was 5'- ATT TAA ATA CTG GGA AAC TTC TTC TAA AGC -3'. A TEV protease linker was incorporated in the design such that the desired protein with only an additional glycine on the amino terminus could be prepared by enzymatic cleavage from MBP. Glycine was chosen to minimize potential effects on protein function due to its small size and neutral charge.

Creation of RNA Probe for DAZL Binding

Venables et al. determined an RNA consensus sequence preferred by DAZL in mice[84]. A series of oligonucleotides were designed based on three long

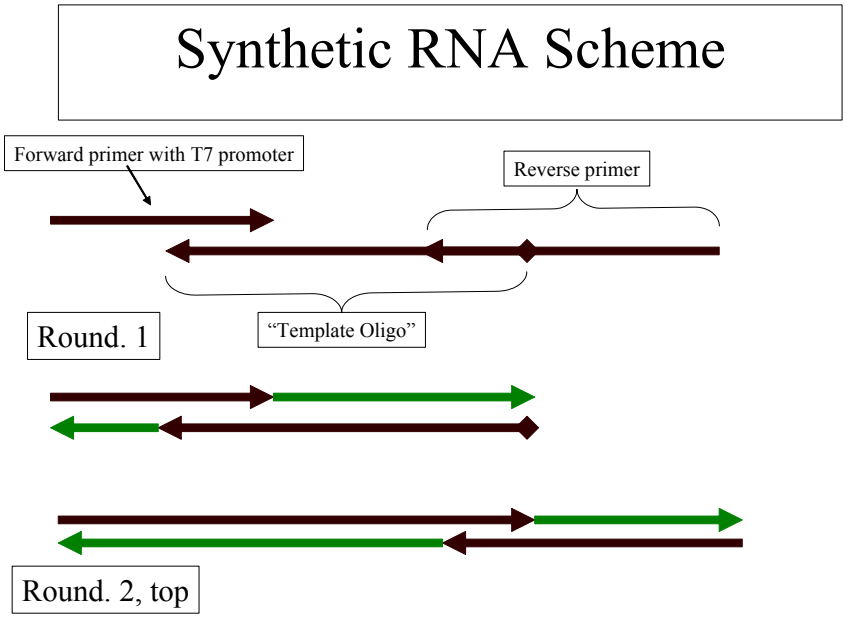


Figure 13. Schematic of Synthetic RNA Scheme. The brown arrows indicate primers that were ordered. The green arrows indicate the newly-synthesized DNA. In round one, only the forward primer with the T7 promoter binds with the "template" oligo where extension progresses in two directions. In round two, the reverse primer is able to bind the extension primer. Polymerization proceeds in both directions giving the full, final product.

oligonucleotides (ordered from IDT) that contained the repeat region $GU_7GU_{10}GU_{10}GU_7$. The forward primer, T7Univ, contained the sequence 5'-CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT -3'. The reverse primer, RevUniv_2 contained the sequence 5'- CCC GAC ACC CGC GGA TCC ATG GGC ACT ATT TAT ATC AAC -3'. Finally, the "template" oligo, CDC25C_2,

contained the sequence 5'- TGG GCA CTA TTT ATA TCA ACA AAA AAA CAA AAA AAA AAC AAA AAA AAA ACA AAA AAA CAA TGT CGT TGG TGG CCC - 3'. The 100 µl PCR reaction consisted of 52 µl ddH_2O, 10 µl of 10X PCR Buffer, 10 µl of 25 mM $MgCl_2$, 8 µl of 2.5 mM each dNTP mix, 5 µl of the 0.1 µg/µl T7Univ primer, 9 µl of the 0.1 µg/µl RevUniv_2 primer, 1 µl of 5 U/µl AmpliTaq Gold, and 5 µl of the 0.01 µg/µl CDC25C_2 "template" oligo. The PCR program was run on a Stratagene Robocycler at 95°C for 20 minutes, 35 cycles of 95°C for 15 seconds, 60°C for 30 seconds, and 72°C for 5 minutes. After PCR was performed to concatenate the sequences, the PCR product was cloned into pCR2.1 using the quick cloning method mentioned previously, and the sequence verified. Figure 13 depicts the basic scheme for creation of the probe from three oligonucleotides (Appendix 3). The 5' end of the sequence contains a T7 promoter for *in vitro* transcription in the presence of ^{32}P-UTP. The resulting PCR product was 121 bp and the radiolabeled probe was 95 bp.

In vitro transcription was performed by incubating the DNA generated through PCR with T7 polymerase, nucleotides, and a radiolabeled nucleotide. *In vitro* transcription was carried out using Ambion's MaxiScript T7 Kit (Cat# 1314, Austin, TX). The reaction conditions contained 5 µl of the PCR-generated CDC25C template, 2 µl of 10X transcription buffer (specific contents not disclosed by the company), 1 µl of 10 mM ATP, 1 µl of 10mM CTP, 1µl of 10mM GTP, 2 µl of 50 µM UTP, 3.5 µl of nuclease-free water, 2.5 µl of [α-^{32}P] UTP (Amersham Pharmacia, Cat# PB20383, 20 mCi/ml, 800 Ci/mmol, 250 µCi total), and 2 µl (15 units/µl) T7 RNA Polymerase. The

reaction was incubated at 15°C for 3 hours whereupon 1 µl of (2 units/µl) DNase I was added. Another incubation period at 37°C for 15 min. followed.

A control template found in the Ambion MaxiScript T7 Kit (Austin, TX) was used in addition to other probes. The mouse antisense β-actin (pTRI-Actin-Mouse) probe was labeled as per manufacturer's instructions and treated the same way as the other probes used in this study.

Specific activity was determined by comparing the percentage of TCA-precipitated material with information in Ambion's *in vitro* Maxi-Script kit instructions (Austin, TX). Briefly, TCA precipitates RNA enhanced by the presence of a carrier such as sheared salmon sperm DNA. The solution is poured through a filter under vacuum and the ratio of what is counted on the filter is compared to the input RNA.

After radiolabeling the RNA it was necessary to determine the specific activity of the probe. Radiolabel incorporation was assessed through TCA (trichloroacetic acid) precipitation. One hundred fifty microliters from a solution of 1 mg/ml Sheared Salmon Sperm DNA (Ambion, Austin, TX Cat#9680) was pipetted into a 1.5 ml centrifuge tube. To this solution, 2µl of the transcription reaction was added and mixed well with pipetting. 50µl of the mixed DNA and radiolabeled RNA was pipetted onto a Whatman GF/F glass filter. Another 50µl of the mixture was pipetted into a borosilicate tube to which 2ml of 4°C 10% TCA was added. This mixture was allowed to incubate on ice for 10 minutes. After 10 minutes the solution was vortexed and poured through a Whatman GF/F glass filter under minimal vacuum in a filter setup. The tube was rinsed 2X with 1 ml of 4°C 10% TCA with vortexing and pouring through the GF/F filter.

Finally the tube was rinsed with 5ml of 4°C 95% ethanol and applied to the filter under minimal vacuum.

Each of the filters was dried for 30 minutes under a sunlamp to evaporate all of the ethanol remaining on the filters because ethanol is a known quencher of beta particles in scintillation fluid.

After the filters had dried, they were placed in glass vials with 5 ml Ecoscint XR (National Diagnostics) and vortexed. They were read in a Beckman 6500 scintillation counter for 5 minutes each and the counts were documented.

Specific Activity Calculations

The following is an example of how to calculate specific activity from the synthesis of a probe found later in this document. The calculations stay the same, but the number and complement of specific nucleotides are adjusted based on the probe generated. The amount of precipitated material was 227,055.00 CPM whereas the total signal was 1,419,675.20 CPM. Therefore 227,055.00 / 1,419,675 = 0.16. So roughly 16% of the material that was synthesized was incorporated into precipitated RNA. The next step was to calculate the number of moles of [α-^{32}P]UTP in the reaction. To do this the number of millicuries (mCi) of UTP in the reaction is calculated. In the reaction, 3 µl of 10 mCi/ml UTP was added into the mix. The equation for the calculation follows:

$$\frac{0.003\text{ml} * 10\text{mCi}}{1\text{ml}} = 0.03\text{mCi}$$

.

The next step was to calculate the number of millimoles of [α-^{32}P] UTP in the reaction.

$$\frac{0.03\text{mCi} \times 1\text{mmol}}{3000\text{Ci}} \times \frac{1\text{Ci}}{1000\text{mCi}} = 1 \times 10^{-8}\,\text{mmol} = 10.0\,\text{pmoles}$$

In addition, the moles of unlabeled UTP in the reaction was calculated as:

$$1\upsilon l \times \frac{10{,}000\upsilon\text{mol}}{1000\text{ml}} \times \frac{1\text{ml}}{1000\upsilon l} = 0.01\upsilon\text{mol} = 10{,}000\,\text{pmol}$$

It follows that the total amount of UTP in the reaction was 10.0 pmoles labeled UTP + 10,000 pmoles unlabeled UTP = 10,010 pmoles total UTP in the reaction. The concentration of total UTP in the reaction was calculated as 10,010 pmol / 20 µl = 500 pmol/µl = 500 µM. Next, total UTP incorporated into the RNA was calculated by multiplying the total number of pmoles of total UTP in the reaction by the fraction incorporated into RNA: 10,010 pmol x 0.159934 = 1,601 pmol

The number of moles of UTP incorporated into the RNA was used to calculate the mass of RNA synthesized. The transcript from NM_014814 contains 1275 bases which consist of 390 A residues, 250 C residues, 307 G residues, and 328 U residues. The fraction of the transcript given by each nucleotide is calculated in Table 8.

Table 8. Nucleotide Base Count for NM_014814.

Base	Count	Fraction of Total Transcript
A	390	0.306
C	250	0.196
G	307	0.241
U	328	0.257
Total	1275	1.000

Since the amount of UTP incorporated was 1,601 pmoles, the other residues can be calculated by first dividing the pmoles of UTP by its complementary fraction to give the total amount synthesized. 1,601 / 0.257 = 6230 pmoles.

From this total amount synthesized, the number of picomoles of the other nucleotides present in the transcript can be calculated. From this, the amount synthesized in grams is derived utilizing dimensional analysis using residue A as an example. This was done for all of the nucleotides and the results were added to determine total amount of RNA synthesized in grams.

$$1910\text{pmoles} \times \frac{1\text{mole}}{1000\text{mmoles}} \times \frac{1\text{mmole}}{1000\text{nmoles}} \times \frac{1\text{nmole}}{1000\text{pmoles}} \times \frac{329.208\text{g}}{1\text{mole}} = 6.29\text{E} - 07\text{g}$$

Table 9. RNA Synthesis Calculations.

Nucleotide	Base Count	Fraction	Pmoles Total RNA	Pmoles Nucleotide	MW	Amt. Synthesized (g)
A	390	0.306	6,223.18	1910	329.208	6.29E-07
C	250	0.196	6,223.18	1220	305.183	3.72E-07
G	307	0.241	6,223.18	1500	345.207	5.18E-07
U	328	0.257		1,601	306.168	4.90E-07
Totals	1275	1.000		6230		20.09E-06

The resultant final amount of RNA synthesized in nanograms was 20.09 ng (Table 9). The number of counts per minute incorporated into the product was

determined from the number of µl represented in the reaction as 2 µl / (2 µl + 150 µl) = 0.7 µl counted on the filter containing the precipitate. From this the CPM/µl was determined to be the same as CPM of the filter divided by the number of µl counted, i.e., 227,055 CPM / 0.7 = 300,000 CPM / µl. The volume of the transcription reaction before TCA precipitation was 22µl so 22 µl x 300,000 CPM / µl = 6,600,000 CPM. Finally, specific activity can be calculated by dividing the total number of counts per minute by the number of nanograms of RNA synthesized, 6,600,000 CPM / 20.09 ng = 328,522 CPM/ng, or by dividing the total number of counts per minute by the number of pmoles synthesized 6,600,000 CPM / 6,230 pmoles = 1,059 CPM / pmole.

RNase Detection

RNase contamination in laboratory reagents was detected using a commercially-available kit from Ambion (RNaseAlert Lab Test Kit, Cat# 1964, Austin, TX). Forty five microliters of the solution to be tested was added to sample substrate and buffer which was included in the kit. The mixture was incubated at 37°C for 25 minutes. The results were read in a fluorometer where the substrate was excited at 490 nm and read at 520 nm.

Gel Shift with CDC25C

A gel shift was performed to test whether or not DAZL bound CDC25C RNA construct. The binding reaction consisted of 2 µl of water, 5 µl of binding buffer (final: 20mM Tris-HCl, pH 7.6, 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol, and 100 µg/ml BSA), 1 µl of SUPERaseIn (Ambion, Cat# 2694, Austin, TX), 1 µl of RNA construct (710 pM final), and 1 µl of the DAZL protein construct (a 2-fold reduction serial dilution from 8 µM to 7.8 nM). The reactions were incubated at 36°C for 15 minutes and electrophoresed on a pre-run (80 volts for 15 minutes) 7.5% Tris-HCl BioRad Criterion Precast Gel (Cat# 345-0005) for 1 hour and 40 minutes at 80 volts in tris glycine buffer. The gel was dried between cellulose and mylar and detected on a phosphorimager screen for three hours at room temperature. MBP was purchased through NEB (Ipswich, MA), Cat # 800-44S.

Nitrocellulose Filter Binding

1- 4" X 5.25" piece of nitrocellulose paper (Schleicher & Schuell Protran Bioscience Pure Nitrocellulose Transfer and Immobilization Membrane, BA83, Pore Size 0.2 µm, Cat# 10402488, Keene, NH) was incubated at room temperature in 1X Binding Buffer (20 mM Tris, pH 7.4, 50 mM KCl, 0.75 mM $MgCl_2$, and 4.5 % Glycerol) for 30 minutes prior to the experiment. 20 µl binding reactions consisting of 1 pmol of radiolabeled RNA (50 nM final), 10 µl of 2X binding buffer (same as above for the nitrocellulose), and 5 µl of protein diluted appropriately for series. The binding reactions were performed in PCR tubes on ice for 30 minutes and applied to the

membrane through the filter and chased with 50 µl 1X binding buffer. A template was used to cut out the proper portions of the filter and the dots were placed in scintillation vials and counted 5 minutes in a Beckman LS6500 scintillation counter.

Gradient Centrifugation

A 5-30% glycerol gradient in binding buffer was created using a 12 ml gradient maker based on methods utilized by Dember et al[85]. The gradient consisted of two containers, one containing a 5% glycerol solution in binding buffer and the other containing a 30% solution of glycerol in binding buffer. The mixture was fed through a small tube into clear polycarbonate centrifuge tubes from Beckman that were balanced within 1 mg of each other with the 5% solution. The binding reactions were applied to the top of the gradients and the samples centrifuged in an ultracentrifuge at 35,000 RPM in a Beckman SW41-Ti rotor for 40 hours at 4°C. After centrifugation the bottom of each tube was pierced with a needle from a fractionator and drops were collected into separate scintillation vials. Five milliliters Ecoscint XR was added to each tube and each tube was counted in a Beckman LS6500. This technique was based on work done by Baumann et al. [86].

Affinity Chromatography Binding Experiments

Proteins used in the binding reaction were expressed in 1 liter of LB with antibiotics from *E. coli* strain ER2508, extracted from the crude lysate after utilizing the french press, and washed from other non-bound proteins from amylose resin according

to New England Biolabs' instructions. Binding reactions for each of the three experiments were as follows. Experiment 1 contained 22 nM final of CDC25C RNA construct in the binding reaction. Columns were pre-treated with 10 µl of 0.1 mg/ml BSA in 1X Binding Buffer (20 mM Tris, pH 7.4, 50 mM KCl, 0.75 mM $MgCl_2$, and 4.5% glycerol. The reaction mixture consisted of 25 µl of resin containing the washed protein of interest, 9 µl of water, 40 µl of 2X binding buffer, 1 µl of 5 mg/ml BSA, and 5 µl of RNA for a total volume of 80 µl. The reaction was allowed to proceed on ice for 20 minutes. In experiment 2, the columns were pre-treated with 10 µl of 0.1 mg/ml BSA in 1X binding buffer with centrifugation for 1 minute at maximum rpm on a tabletop centrifuge. The reaction mixture consisted of 100 µl of resin with attached and washed proteins, 5 µl of RNA (2.2 nM final), and 50µl of 1X binding buffer. The binding reaction was allowed to proceed on ice for 20 minutes. In experiment 3, the Qiagen Small Spin Columns were pre-treated with 100 µl of 100mg/ml BSA in 1X binding buffer. The reaction mixture was 500 µl of resin, 250 µl of 1X binding buffer, and 25 µl of RNA (13.5 nM final). In each case, the binding reactions were applied to Qiagen Small Spin Columns and centrifuged for 1 minute at maximum rpm on a tabletop centrifuge. The flowthrough was applied to 5 ml of Ecoscint XR and counted in a scintillation counter (LS6500, Beckman, Fullerton, CA).

Results

The first step in producing active recombinant DAZL was the use of PCR to amplify the coding region of DAZL from a cDNA library. PCR primers were designed to include a TEV protease cleavage site which allows excision and removal of the MBP (maltose-binding protein) portion of the fusion protein. The template for PCR was a Human Testis cDNA library (Clontech, Mountain View, CA) in the λTriplEx2 phage vector. Results in Figure 14 indicate that the PCR product in the "+" lane contains the coding sequence of DAZL of 930bp.

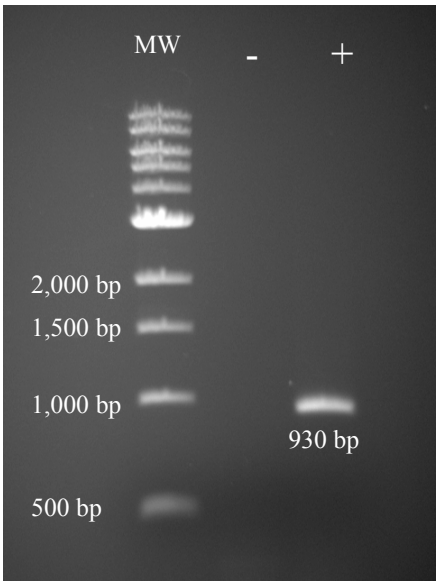


Figure 14. PCR of DAZL from Clontech Phage Library. The gel indicates PCR product of DAZL + TEV from phage library in the + lane at the proper size, the 1 kb MW ladder , and the negative control of water (-).

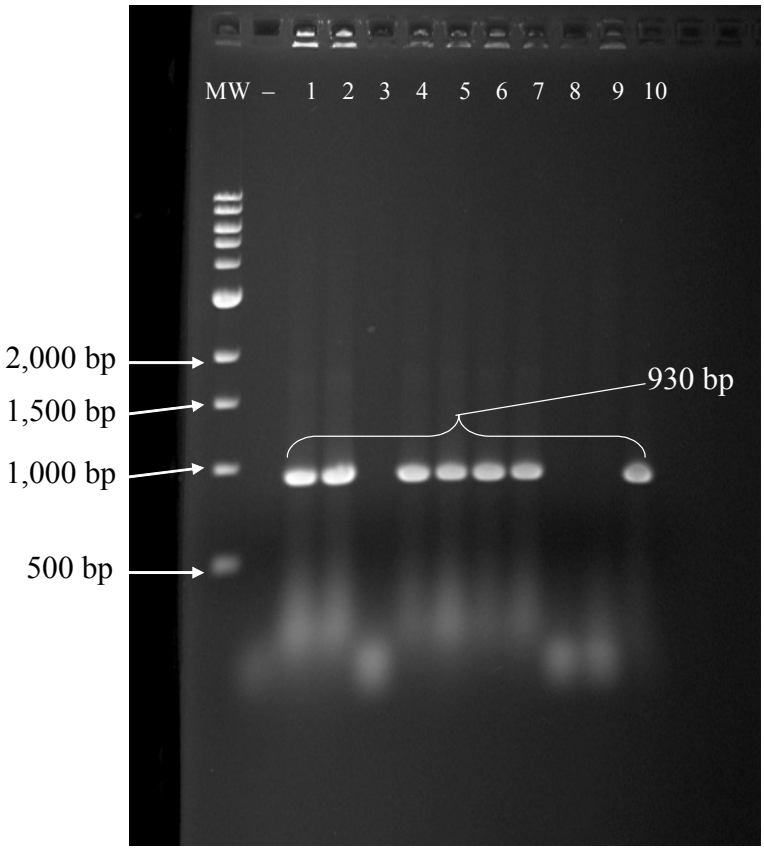


Figure 15. PCR Products from Positive Clones from the pCR2.1 Cloning Reaction. The first lane is a 1kb ladder, The second lane is the negative control of water, and the remaining lanes are PCR reaction products indicating the presence of DAZL of the correct size.

The DAZL PCR product was ligated into the TOPO pCR2.1 vector using the manufacturer's short cloning protocol and transformed into TOP10 cells. The transformed cells were plated onto a selective LB plate and incubated overnight at 37°C and colonies tested with PCR. Figure 15 shows clones 1, 2, 4, 5, 6, 7, and 10 that were positive for the presence of the DAZL.

A miniprep of clone 1 (Figure 15) was used to process the plasmid DNA for sequencing. Modification of the protocol included splitting the cell pellet into two

aliquots and titrating the cell lysis buffer. Serial precipitation and centrifugation were used to assure removal of protein. All samples were applied to 1 Qiatip 100 column

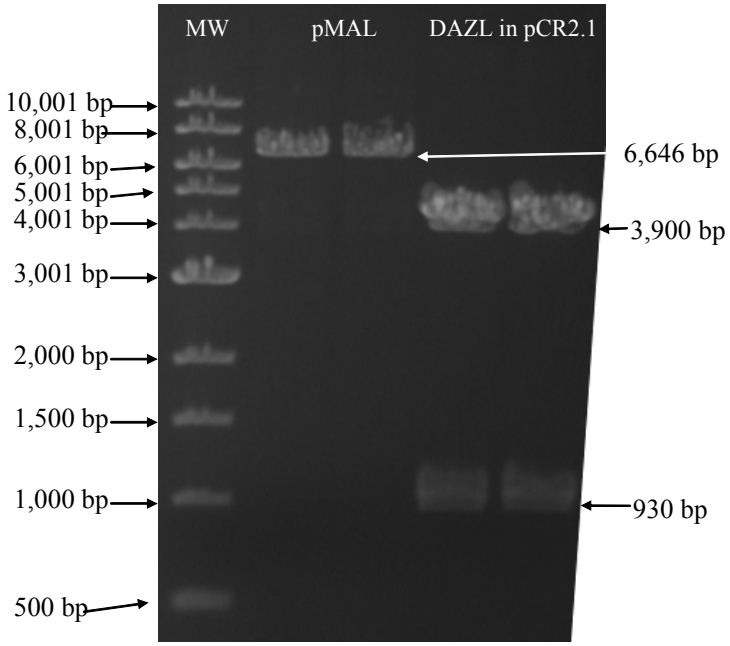


Figure 16. EcoRI Digest of pMAL and pCR2.1. The first lane is the marker, the next two lanes contain the digest of pMAL, and the last two lanes contain the digest of pCR2.1 containing the DAZL fragment.

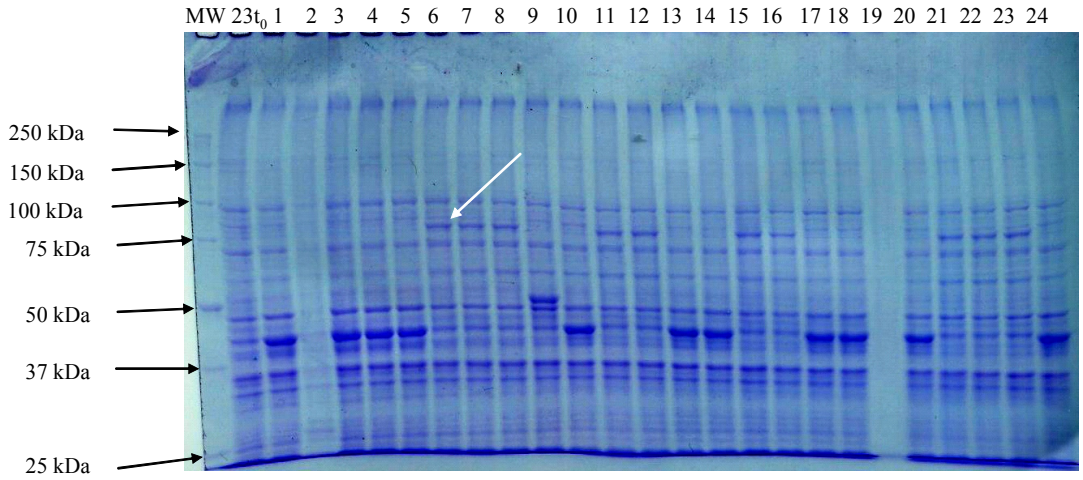


Figure 17. Test Expression of pMAL DAZL Clones. The first lane is the molecular weight standard and the second lane is the t_0 timepoint of clone number 23 used as comparison against clones that induced [this sentence is not complete]. The remaining lanes are proteins expressed by clones used in the assay. The darker bands found below the clones expressing DAZL are MBP. [What about bands above DAZL?] The clones that produced DAZL are 6, 7, 8, 11, 12, 15,16, 21, 22, and 23 and are of the right size (~ 77 kDa).

(Qiagen) and eluted with 150µl of elution buffer. The miniprep DNA was submitted to Lone Star Labs for sequencing and the DNA sequence verified using Accelrys' (San Diego, CA) DSGene Software. The sequence showed 100% identity at the nucleotide level with DAZL in the NCBI database.

After verification of the clone sequence in pCR2.1, the fragment was excised using EcoRI (Figure 16) and ligated into similarly cleaved pMAL vector. One feature of the pMAL multiple cloning site is the addition of MBP at the amino terminus of the protein which facilitates affinity purification on amylose beads. Initially, the relaxed

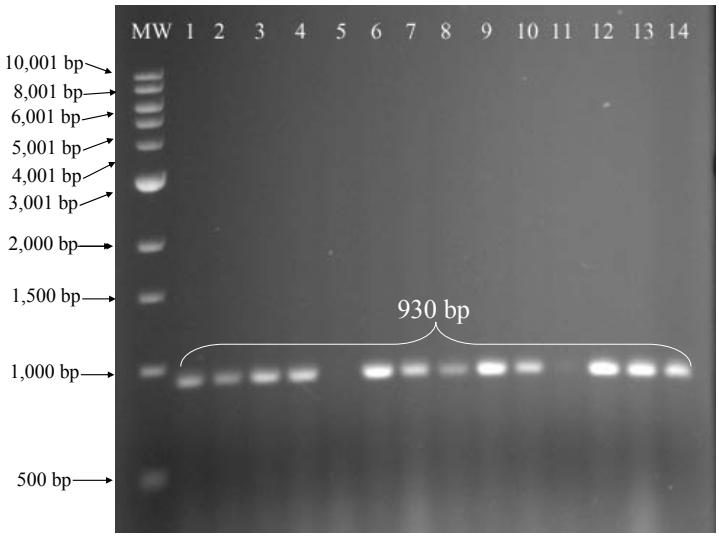


Figure 18. PCR Products from PCR Reaction after Mutagenesis to Fix the DAZL Clone. The PCR product bands are of the correct size at 930 bp.

circle of ligated DNA was introduced into chemically-competent TOP10 cells. To verify the presence of the DAZL inserts, 44 test inductions were performed of which 24 are shown in Figure 17. Clones positive for growth (6, 7, 8, 11, 12, 15, 16, 21, 22, and 23 were grown and saved as freezer stocks. Clone #6 (Figure 17) was arbitrarily selected, grown, and used for a plasmid midiprep, according to manufacturer's instructions (Qiagen), to obtain plasmid DNA. The plasmid DNA was electroporated into protease deficient ER2508 cells to maximize the yield of intact DAZL protein after expression and enrichment.

Sixteen of the resulting MBP-DAZL clones were selected for test induction. One of the positive clones was chosen for the midiprep procedure. The purified plasmid DNA was subsequently sequenced to verify the presence of DAZL and it was found that

the DAZL sequence of this plasmid contained a non wild-type mutation. The most probable cause of the mutation was that the single nucleotide mutation was missed in the

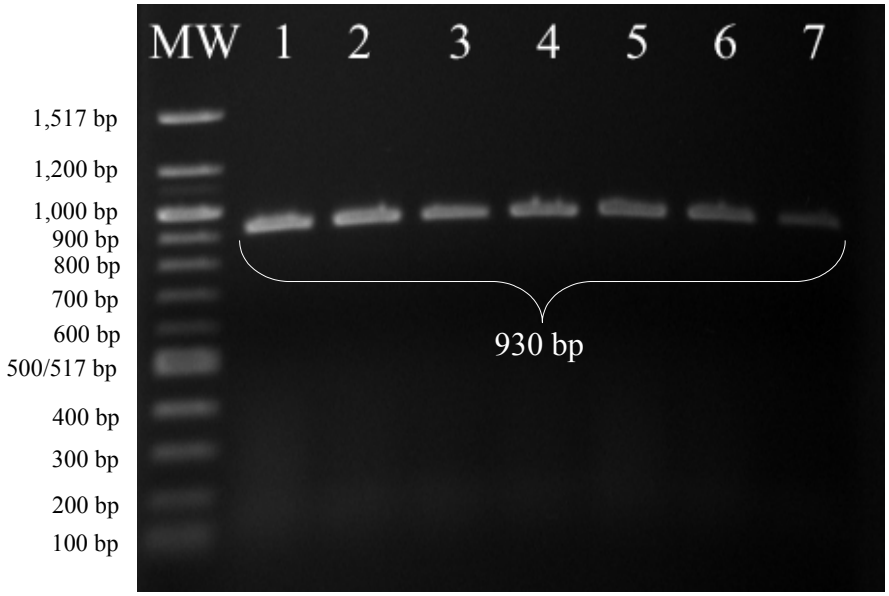


Figure 19. PCR to Determine the Presence of DAZL in ER2508 Cells. All seven clones were positive for the presence of the DAZL insert. The first lane is the molecular weight standard and the rest of the lanes are PCR products of the correct size (930 bp).

original sequence data after the original PCR and passed down to the pMAL insert before being discovered. The mutation was repaired using site-directed mutagenesis from Stratagene's QuikChange Site-Directed Mutagenesis Kit. Clones of the repaired DAZL were tested for the presence of the inserts (Figure 18), sequenced, and transformed into ER2508 cells and verified by PCR (Figure 19). Teng et al. reported[15] two point mutations in the DAZL locus- one inside the RRM of DAZL and one outside of the RRM of DAZL. These mutations were used to change the amino acid

sequence of the DAZL clone and determine the impact of two naturally-occurring mutations on DAZL function. As before, the cells were grown, a midiprep was performed and the resulting plasmid DNA was electroporated into ER2508 cells. During this process, clones were directly verified through sequencing for the presence or absence of the DAZLmut1 (outside of RRM, nucleotide 260) or DAZLmut2 (inside of RRM, nucleotide 286) inserts (Figure 20) and Appendix 1. The PCR contained the original DAZL cloning primers and verified the presence of DAZL.

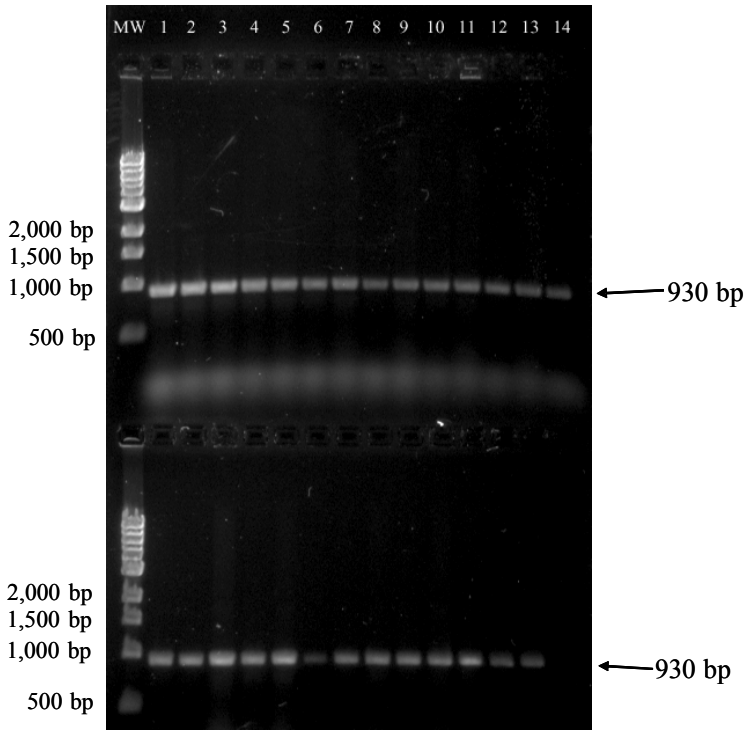


Figure 20. Gel Showing Presence of DAZLmut1 and DAZLmut2. The top row is DAZLmut1 and the bottom row is DAZLmut2. The first lane in both cases is the molecular weight standard. The remaining lanes are separate clones screened for the presence of absence of PCR product.

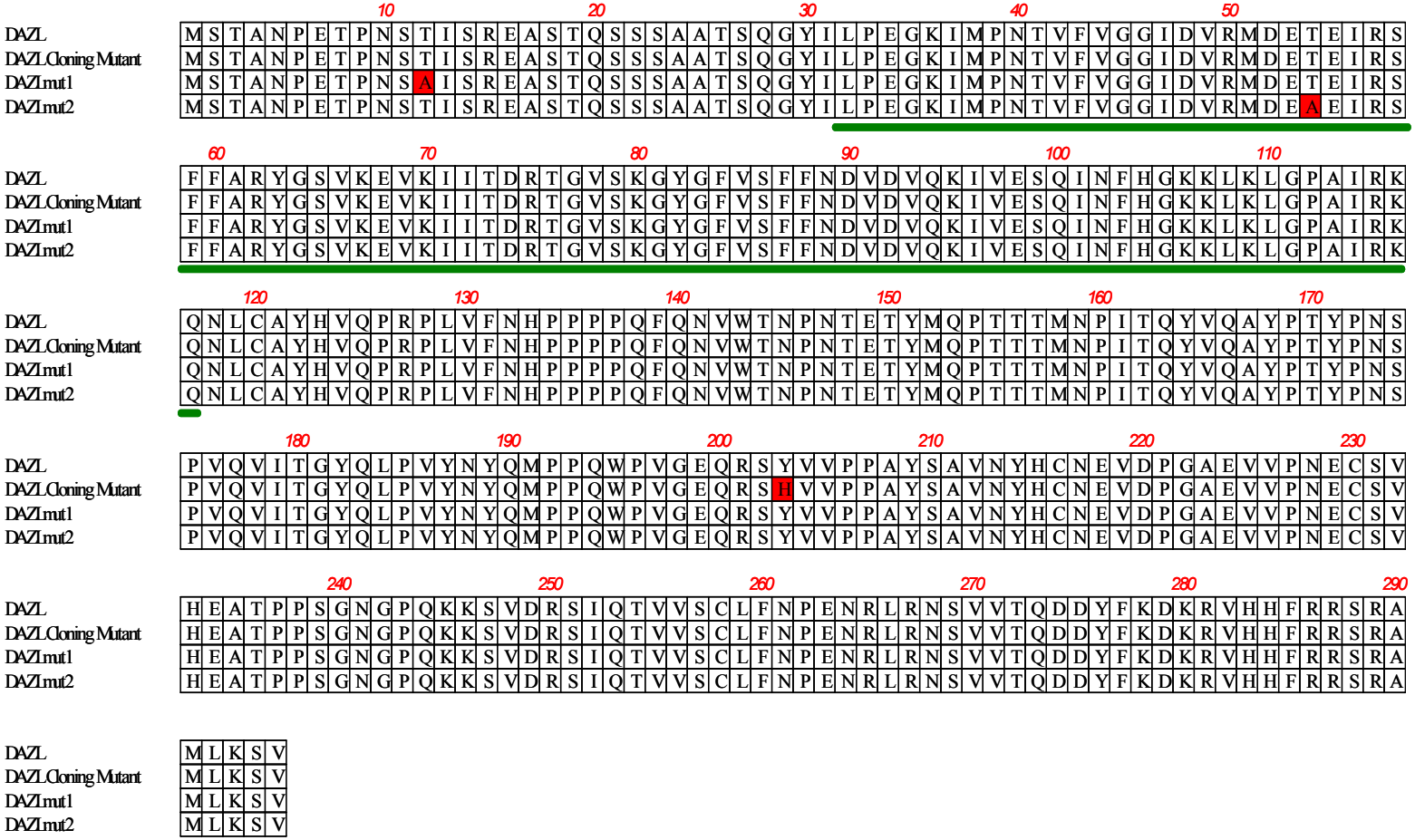


Figure 21. Amino Acid Alignments of DAZL, DAZL Cloning Mutant, DAZLmut1, and DAZLmut2. The red boxes highlight the mutations of interest. The green bar represents the RRM of the protein.

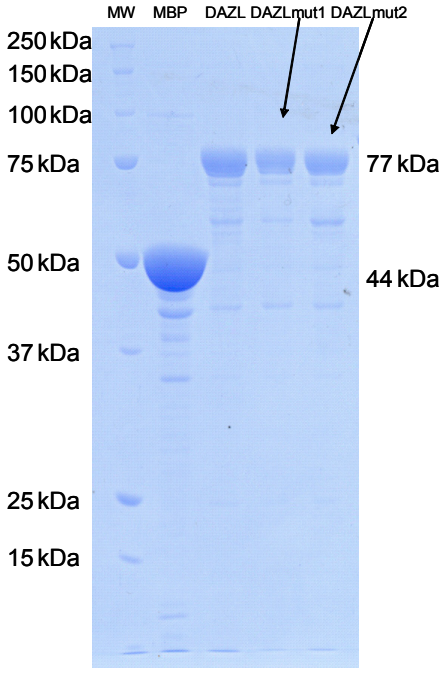


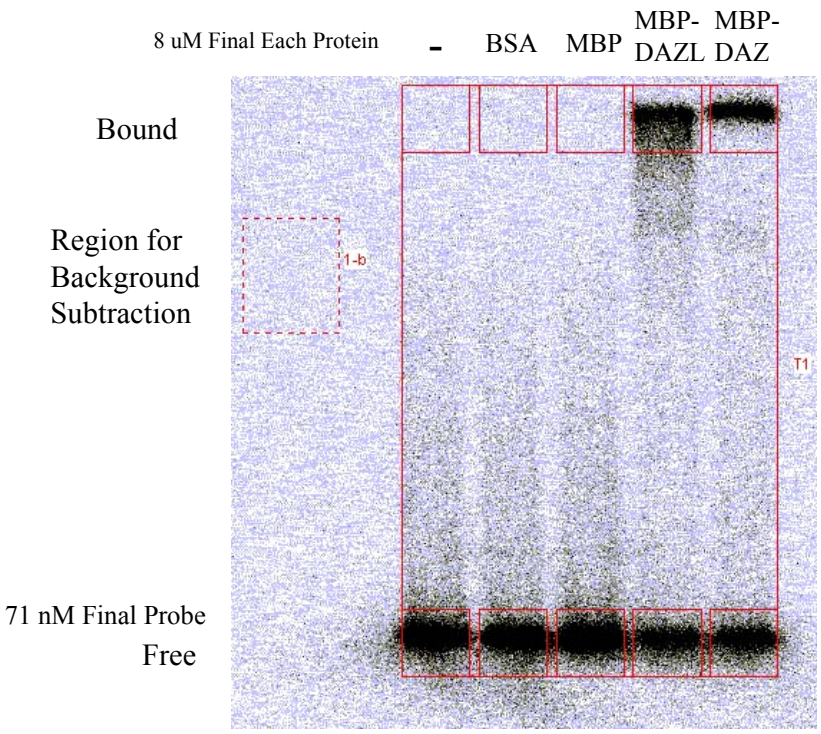
Figure 22. Enrichment of MBP, DAZL, DAZLmut1, and DAZLmut2. A series of washes and the eluted proteins from the amylose resin indicates proteins of the correct sizes expected.

The amino acid sequences of DAZL, DAZL with the cloning mutation, DAZLmut1 and DAZLmut2 are shown in Figure 21. In addition, Figure 22 shows the enriched proteins of MBP, DAZL, DAZLmut1, and DAZLmut2 after binding the proteins to the amylose resin, washing away unbound proteins, and elution of the bound proteins. Of note is the low presence of other proteins. Because there was such a high signal to noise ratio of desired DAZL proteins to contaminants it was decided not to further purify the proteins. The gel was overloaded to increase the signal of contaminants. A visual comparison

demonstrates that among the DAZL preps, the same non-DAZL proteins were present as can be detected by size alone and what remains in one is in the other. A possibility is that there is a small amount of degraded DAZL protein in each of the samples. The size of the DAZL proteins was determined by comparison with the molecular weight marker in comparison to the reported value of 77 kDa (Genbank).

The DAZ2 was also cloned in a manner similar to the cloning of DAZL and was also examined for sequence by alignment with DAZ2 database sequence. This DAZ gene contained a single RRM but lacked repeats reported in the published version of DAZ2. For sequence analysis see Appendix 2.

The function of DAZL was first addressed using an electrophoretic mobility shift assay using an RNA corresponding to a consensus sequence derived from CDC25C. In the first assay, Figure 23, a binding reaction containing a final concentration of 71nM for the CDC25C probe was combined with 8 μM of each protein and applied to a polyacrylamide gel. The results from the initial gel shift indicated that neither BSA nor MBP were able to shift any of the RNA, but both DAZL and DAZ were able to shift roughly 50% of the RNA at a probe concentration of 71 nM.



	- Control	BSA	MBP	DAZL	DAZ
Top	952.30	1,202.10	1,515.90	139,023.00	145,343.10
Bottom	291,512.50	296,771.60	316,984.10	146,454.90	161,362.40
Total	292,464.80	297,973.70	318,500.00	285,477.90	306,705.50
Fraction Shifted	0.33	0.40	0.48	48.70	47.39

Figure 23. Electrophoretic Mobility Shift Assay Control. The background region is labeled with a hatched line and the boxes for reading the densities of the bands are shown as solid. Each protein is indicated above each lane. The data corresponding to the top and bottom boxes are indicated along with the fraction shifted. MBP was purchased from NEB.

Next, a gel shift to estimate the Kd of DAZL for the CDC25C probe was performed (Figure 24). For this assay, a consensus sequence of RNA containing a U-rich sequence was used as before but in a lower concentration. A gel shift analysis was performed with a limiting amount of RNA and increasing amounts of DAZL protein (Figures 24 and 25). From a 2 hour exposure to the phosphorimaging screen it was evident from the calculations that given an input concentration of RNA, DAZL bound

with a Kd of 1.03 µM. When the same procedure was performed after an exposure to the screen for 4 hours (Figure 26), the calculated Kd was 0.43 µM.

Because there were not many data points within the range of the Kd which contributes to inaccuracy in calculating Kd, another gel shift was performed which contained protein concentrations within the transitional range of the curve, Figure 27. From this basic initial analysis it was found that DAZL bound this particular

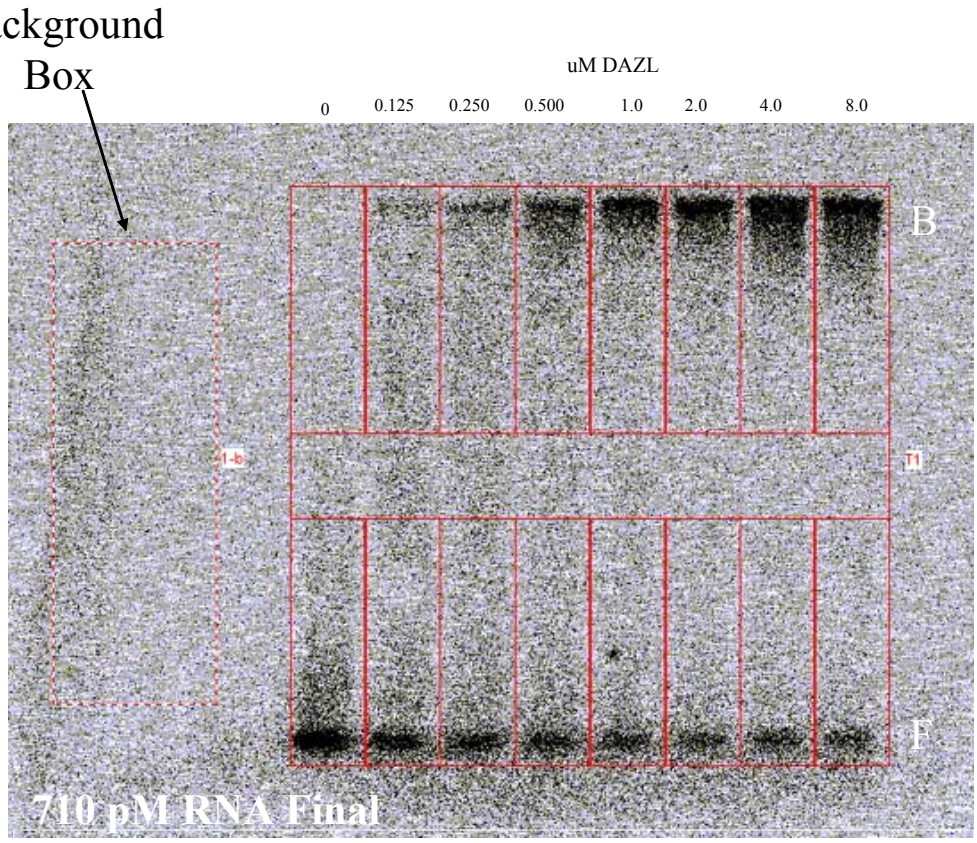


Figure 24. DAZL and CDC25C Gel Shift to Estimate Kd. Input RNA was 710 pM and protein concentration is indicated above each of the lanes. The bound species is on top and the free species of radiolabeled RNA is on the bottom. The background subtraction box for the calculations is shown.

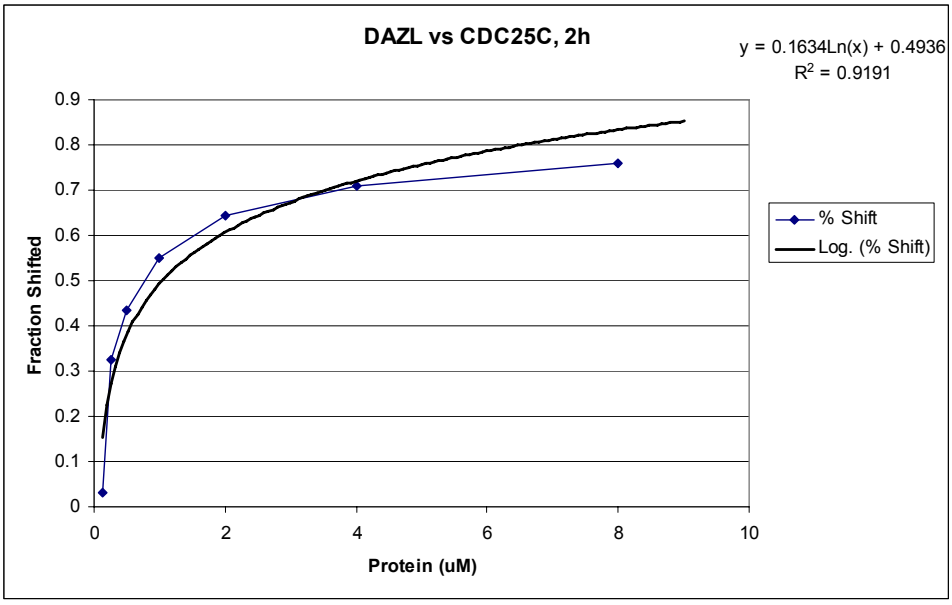


Figure 25. Graph of 2h Exposure to Phosphorimage Screen. The x-axis is the protein concentration and the y-axis is the fraction of radiolabeled probe shifted. Final input RNA concentration is 71 nM. A logarithmic curve was fit to the data and the Kd was calculated to be 1.03 µM.

radiolabeled RNA construct where 50% bound was estimated as 1.5 µM of DAZL protein (Figure 28). This value was calculated by fitting a logarithmic curve to the data points in Figure 27 and calculating the concentration where 50% of the RNA target was shifted. Additionally, in this case the R^2 for the fit line was 0.9774. Table 10 summarizes the data from these three experiments.

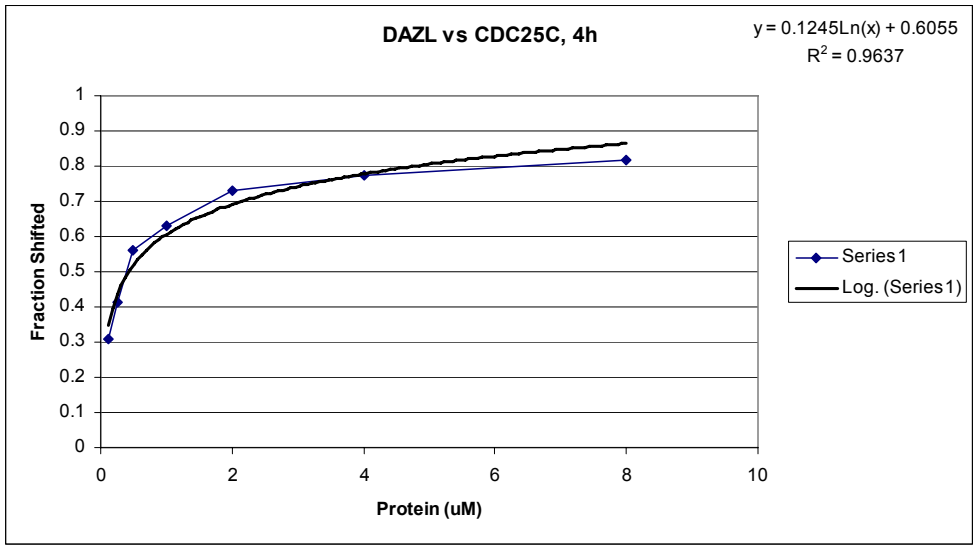


Figure 26. Graph of 4h Exposure to Phosphorimage Screen. The x-axis is protein concentration and the y-axis is the fraction of radiolabeled probe shifted. Final input of RNA is 71 nM. The formula for the logarithmic curve gives a Kd of 0.43 µM.

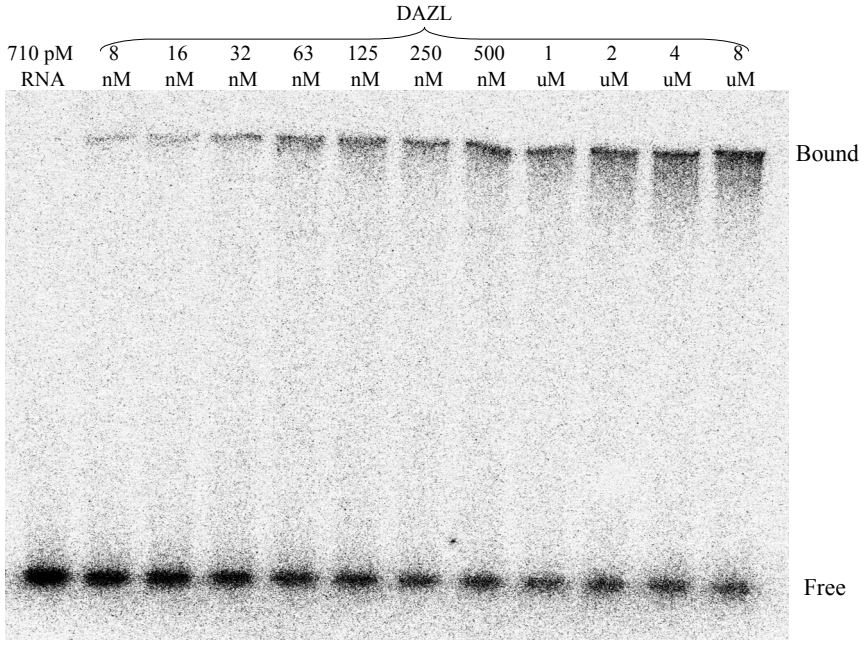


Figure 27. Polyacrylamide Gel Shift of DAZL and CDC25C RNA. The RNA lane contains RNA without any protein. The remainder of the lanes contain increasing amounts of DAZL indicated above.

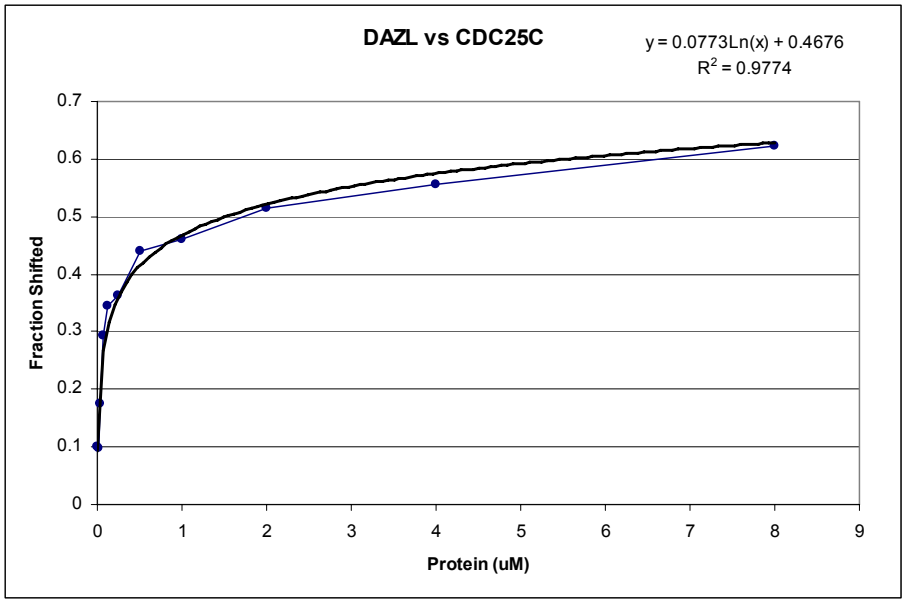


Figure 28. Gel Shift of DAZL and CDC25C RNA. The X axis indicates the concentration of DAZL protein in μM while the Y axis demonstrates the % bound/shifted. As shown in the figure, the static RNA concentration was 710 pM in the reaction. Kd is calculated to be 1.5 μM.

Table 10. Summary of Kd Data.

Figure	R-square	Kd (uM)
24	0.9191	1.03
25	0.9637	0.43
27	0.9774	1.56
	Average	1.01

Because the electrophoretic mobility shift assay is time consuming, the nitrocellulose binding assay was explored in determining Kd of DAZL for RNA because of its relative ease of use and potential for generating large amounts of data quickly. Figure 28 shows the data after binding reactions containing 1 pmole of radiolabeled

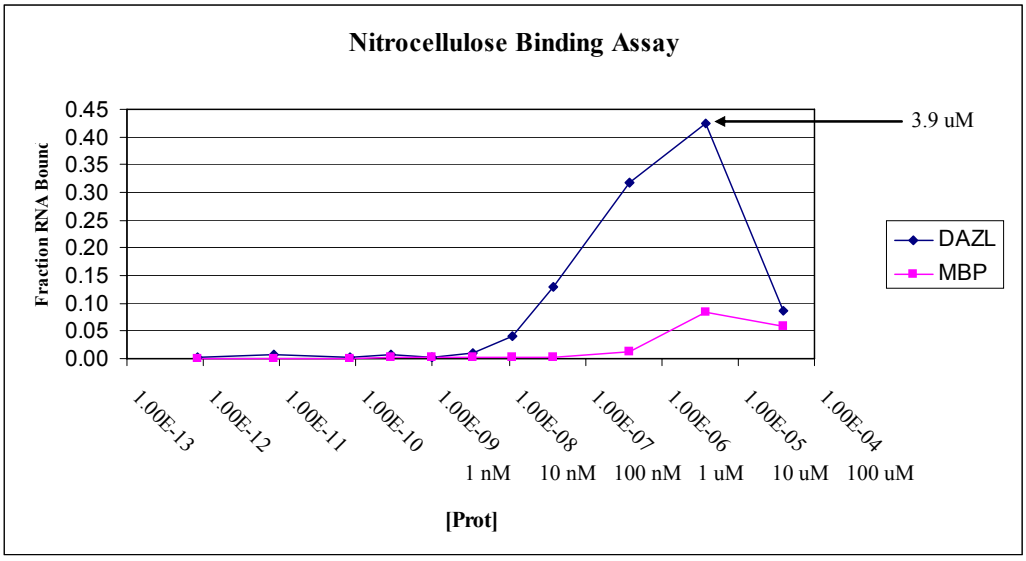


Figure 29. Nitrocellulose Binding Assay. 1 pmole of CDC25C radiolabeled RNA was mixed with increasing amounts of MPB and MBP-DAZL proteins in a binding reaction and applied to a nitrocellulose membrane under vacuum. The MBP used in this assay was generated from an empty vector instead of the commercially-purchased MBP as in Figure 22. Both DAZL and MBP bind RNA although MBP did at a lower extent.

probe and increasing amounts of protein. In this case the MBP used was generated and purified in the same way as the recombinant DAZL. An interesting anomaly occurred in the nitrocellulose binding reaction. Not only did DAZL bind RNA, but so did MBP which had been previously shown not to bind RNA (Figure 23). However, MBP bound RNA much less efficiently than DAZL. Additionally at 15 µM protein, there was a decrease in signal retained for both species and further investigation revealed that the maximum binding capacity of the nitrocellulose was reached in both cases leading to an increased amount of signal passing through the filter.

Because of the anomaly indicated in Figure 29, a 5-30% glycerol gradient was implemented to help visualize the binding of DAZL, and to a lesser extent similarly

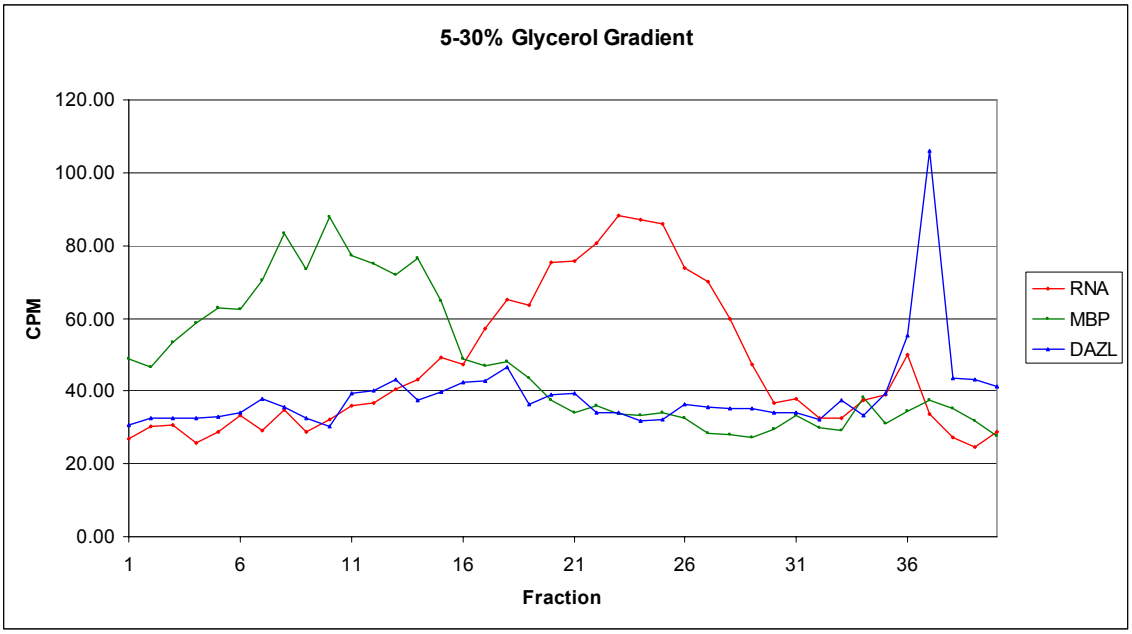


Figure 30. 5-30% Glycerol Gradient and CDC25C. Fractions were taken from the bottom and counted in a scintillation counter. The RNA line shows a fairly wide range over its migration. Upon the addition of MBP, the RNA shifts down with a final pattern which is the same as the RNA by itself. Upon addition of DAZL, a higher less dense band forms which appears over a lower number of fractions.

prepared MBP, to the synthetic target RNA. The binding reaction was applied to the gradient and centrifuged. Fractions of the gradient were read in a scintillation counter and plotted as CPM vs fraction number, Figure 30. The results showed that both DAZL and MBP shifted RNA but to different places in each of the gradients. DAZL shifted the RNA higher in the gradient where the solution is less dense in a narrower band while MBP shifted the RNA to a more dense, albeit wider and possibly non-specific, place in the gradient.

To address the potential RNA targets for DAZL, a binding method involving the use of affinity chromatography with immobilized DAZL was developed. The idea behind this assay was to use amylose affinity resin to separate DAZL-MBP-RNA

complexes from unbound RNA where DAZL-MBP, or any given protein construct with its radiolabeled target, were retained on the column. Because there was concern about changes in binding that might occur related to processing of recombinant proteins, only the target RNA for DAZL binding was radiolabeled and the signal to be counted was that which flowed through the column to minimize steps needed to recover material from the column. Any RNA remaining on the column was calculated as the difference between that added and the amount that flowed through the column. Before any binding reactions in affinity chromatography were to take place, a survey of possible RNase contamination was made. Double-distilled water, reverse osmosis water, and the buffers

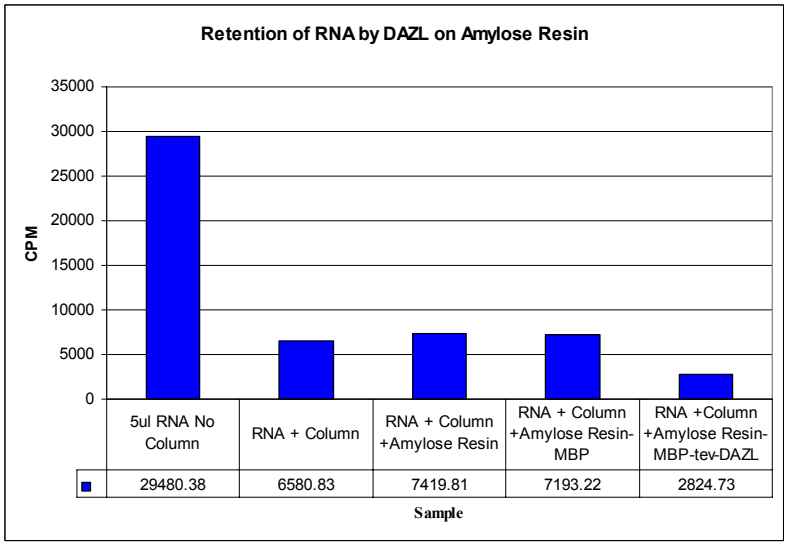


Figure 31. Retention of RNA by Affinity Chromatography System. The first column is input RNA with no treatment. The remaining treatments are indicated under each column and the values generated from scintillation counting are shown below each column. The numbersoc columns in the signal counted in the flowthrough for each of the experiments.

used in the binding reactions were found to be free of RNase contamination using a commercially-available kit from Ambion.

The system that was used for this assay was also assessed as to retention of RNA (Figure 31). The results showed that nearly 80% of the signal was retained non-specifically by this system. Interestingly, when comparing the constructs among themselves, the construct containing DAZL retained more RNA than MBP alone which validated this approach for finding potential binding targets of DAZL (Figure 32). To use this affinity chromatography in a binding assay comparing different protein constructs, MBP, MBP-DAZL, MBP-DAZLmut1, and MBP-DAZLmut2 were incubated with *in vitro*-transcribed RNA for 15 minutes on ice, centrifuged to remove unbound RNA, and the radioactively labeled, unbound RNA counted in the flowthrough. The RNA bound to the resin was also counted by subtracting flow-through activity from the input RNA activity. Figure 33 illustrates that DAZL and both mutant forms of DAZL bound more RNA than MBP alone on the resin. However, the difference between the DAZL forms was not significant ($p > 0.129$) using ANOVA with Tukey's post-hoc test.

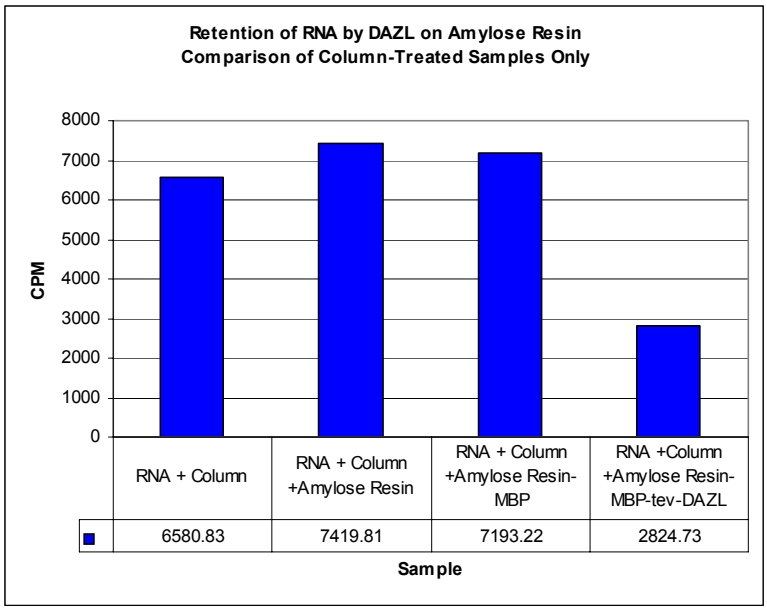


Figure 32. Retention of RNA in Column Treated Samples. When the data from the column-treated only samples from Figure 30 are compared, it is evident that DAZL-containing resin bound more RNA than the resin only containing MBP. The numbers below each of the columns indicate the CPM flowing through the column and counted in a scintillation counter.

arcsine modification of the data was not needed. Variation among experiments was observed, but the ability of DAZL to retain more RNA was established and the DAZL fusion proteins were functional in RNA binding activity. As a control for nonspecific binding, a mouse antisense β-actin RNA binding target was generated and used in another affinity chromatography binding assay (Figure 34). The results from this experiment indicated that the DAZL proteins and their mutants possess an enhanced affinity for the CDC25C RNA binding target when compared with the mouse non-specific housekeeping gene (β-actin).

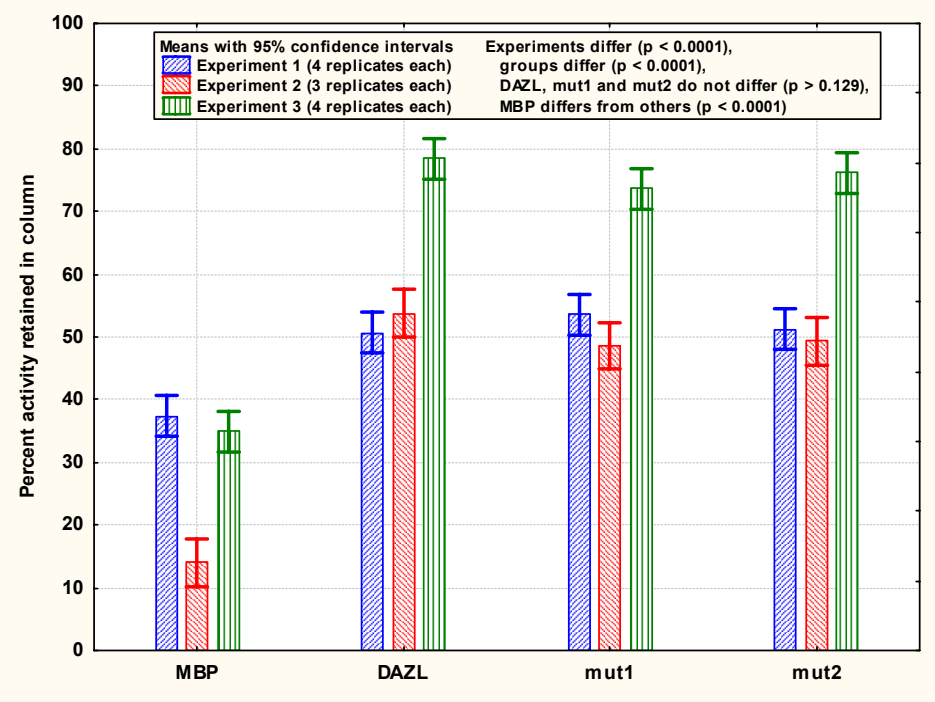


Figure 33. DAZL and Its Mutant Forms Bound More RNA Than MBP Alone. However there was no difference in binding among the different clones. Experiment 1 contained 22 µM RNA, 25µl protein-bound resin, and the columns were pre-treated with 10µl 0.1mg/ml BSA. Experiment 2 contained 2.2 µM RNA, 100µl protein-bound resin, and the columns were pre-treated with 10 µl 0.1mg/ml BSA. Experiment 3 contained 13.5 nM RNA, 500µl resin, and the columns were pre-treated with 100µl of 100 mg/ml BSA.

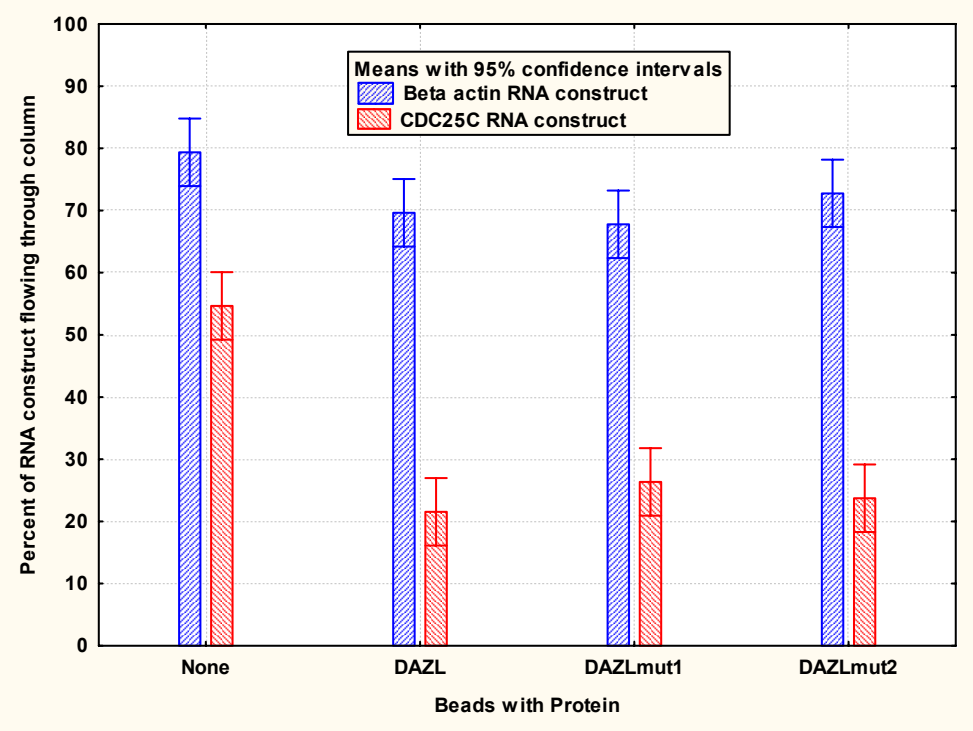


Figure 34. β-actin Control Showing DAZL Binds CDC25C and Not β-actin. The affinity chromatography data show that DAZL and its mutants bind the CDC25C construct in a manner which is more specific than the binding to the mouse antisense control.

Discussion

The first step in examining the function of DAZL was the production of an active recombinant protein. The straightforward sequence of events outlined in the text explaining the cloning procedures belies the difficulty encountered during any cloning process. For example, in the initial attempts the DAZL protein could not be purified as an active, intact form using a common vector that included a His tag. To address this problem a different system was used with MBP as a fusion protein. Other researchers expressing DAZ proteins have used other fusion proteins such as glutathione S-transferase (GST) tags[78]. At Texas A&M University, a popular fusion tag for purification is the MBP tag because of the increased solubility MBP provides and the elution with a dilute maltose solution. In addition, resultant multiple cloning steps into the pMAL vector took extensive effort. pMAL was selected to maintain solubility of the fusion proteins. When DAZ proteins were expressed in the *E. coli*, there was a tendency for an increase in the doubling time of induced cells. To combat possible degradation of the recombinant proteins, growth of all cultures destined for expression were grown at 28°C in a protease deficient cell line (ER2508) designed for protein production. In another series of experiments, Fox et al. used the pET29 vector from Novagen for cloning. In their assay, the purification tag included a 6-His tag[79] which required the use of denaturing conditions to purify the protein followed by dialyzing the protein against a suitable buffer in which to store and use the protein. The problem with this methodology is that there are risks associated with denaturing and refolding the protein.

In this study an MBP tag was used to avoid the harsh conditions. In addition, while the cloned DAZL fragment contained the entire coding region of DAZL (NM_001351.2, nucleotides 295-1182) used by Fox et al., these authors did not indicate whether it was exactly the same as the accession sequence at NCBI [79]. Additionally it became evident that the conditions used to enrich for the desired protein of interest also purified proteins that bound RNA (Figures 22 and 28). This is the most likely hypothesis that can explain why there is no RNA binding of MBP when the protein used in the experiment is purchased commercially as it is in Figure 23. However, when MBP is enriched using the same methodology and the same cell line as what is used to generate the other protein constructs, that "MBP" seems to shift RNA. However, Figure 30 shows that DAZL shifts the RNA to a different place than the MBP construct. Additionally it takes more of the MBP construct to bind to the same amount of RNA than DAZL. Therefore it is likely that whatever protein is present is binding in a relatively non-specific manner.

In the present study, human DAZ and DAZL and unique variants of DAZL representing mutated forms associated with infertility in Taiwanese populations were produced. Another mutated form not associated with infertility was produced to serve as a negative control. The sizes and sequences of these molecules have been verified. Other researchers have produced data describing DAZ interactions[17, 38, 80, 84, 87]. However, most of the descriptions have been based on yeast two hybrid systems or experiments with other species such as in the mouse or *C. elegans*. The human DAZ and DAZL proteins produced in the present study provide a more direct and simpler

approach to determine interactions of the DAZ proteins with other molecules and to explore the kinetics associated with DAZ function.

Up to this point there had been little data determining quantitative binding of the DAZ proteins to RNA. Jiao et al. did examine RNA targets of binding to cloned murine DAZL, but the study was limited to only one isoform.[88]. Venables et al. also performed an RNA association study where cloned murine DAZL was used to elucidate a consensus sequence to which murine DAZL bound. A mutant form of the RRM of murine DAZL exhibiting four point mutations in the RRM was used in their binding assay. None of these investigators reported estimates of Kd for binding activity demonstrated in their systems. The current study assayed not only the DAZL protein but also population-derived mutants. The combination of three different electrophoretic mobility shift assays revealed that DAZL binds RNA with a Kd of 1.0 μM. A decent binding value for an RNA binding protein has been reported as in the 5 nM range with +/- an order of magnitude[82]. Therefore the binding of DAZL to the CDC25C RNA seems less than expected. However, later experiments show that CDC25C may not be a preferred target for the DAZL protein. However, the electromobility shift assay demonstrates definite binding of DAZL for target RNA.

Additionally, the affinity chromatography binding reaction served as an alternate measure for binding after cloning and expression of the DAZL proteins and ultimately showed that all forms of DAZL bound to a synthetic RNA target. Once again it was shown that the affinity chromatography system and possibly a factor or factors co-purified with MBP also bound RNA but not as well as DAZL. In addition, it was shown

through the same binding assay that DAZL preferentially bound the CDC25C RNA target when compared with a general housekeeping β-actin RNA. The column-binding assay also served as a starting point for experiments involved in determining the RNA targets of DAZL.

These experiments ultimately showed that cloned and expressed DAZL, DAZ , and mutant forms of DAZL can bind RNA in a variety of methods such as gel shift, nitrocellulose, or affinity chromatography. Because the control protein MBP copurifies the same proteins as DAZL, MBP can be used as a control for further binding experiments.

By combining the methods of the electrophoretic mobility shift assay and affinity chromatography, experiments can be performed to probe the targets for DAZL and DAZ binding as well as determining the effects of mutations of DAZL and DAZ on the binding efficiencies of each of the proteins for their RNA targets.

CHAPTER IV

FUNCTION OF DAZL

Introduction

Many investigations regarding DAZL or DAZ rely on the presence or absence of genes or modifications of those genes and correlate the results with infertility. However, an underlying question remains. What is the function of the DAZ and DAZL proteins? It is known that DAZL proteins bind RNA and a list of RNAs for both murine DAZL and human DAZL have been described [79, 88, 89]. Disappointingly few references describe cloning of human DAZL or DAZ for eventual expression as proteins for an eventual focus on binding activity. The binding potential of human DAZL and DAZ is based on the presence of an RRM (RNA recognition motif) sequence. To date, procedures such as SELEX (Systematic Evolution of Ligands by Exponential Enrichment), tri-hybrid screening, and affinity purification have been utilized in finding the targets for DAZL binding in mouse and human [79, 84, 88]. There are points to keep in mind when considering such articles. Methods such as SELEX rely on serial amplification of RNA targets and thereby artificially produce enriched numbers of targets to which the protein binds. These methods may prove useful for identifying consensus sequences, but may not yield the complete picture concerning the targets of DAZL binding. In addition, controls in these experiments involving affinity purification of fusion proteins of DAZL did not include the fusion tag in a separate binding reaction as comparison. In one study by Jiao et al. the control protein was a non-specific RNA-

binding protein [88]. A comparison was made between targets selected by the control protein versus the DAZL protein. A possible weakness with this study is that a valid target of DAZL may be discounted due to selection by the control protein. In addition, the same study used total testis extract in its binding assay in hopes that other binding cofactors found within the complement of the extract would yield the true targets of DAZL. In the mouse, DAZL is found expressed in spermatogonial cells from mitosis to the beginning of meiosis (Figure 4)[1]. If a total extract is generated from seminiferous tubules progressing through many different stages of spermatogenesis it may be possible that inhibitors will be included. In addition, in humans, DAZL is found expressed throughout spermatogenesis and DAZ is found expressed from the beginning of mitosis through the beginning of differentiation (as in the mouse)[1] This factor also suggests that caution be used when generalizing on the use of targets of DAZL or DAZ binding in the mouse to predict targets in the human. Furthermore, in Jiao et al.'s study, when the targets of DAZL were bound to the proteins, the bound RNA was purified through multistep procedures where loss of signal becomes a greater issue. After the RNAs were isolated, cDNA was synthesized, cloned into vectors, and analyzed separately in a time-consuming and expensive process.

The method outlined in this study utilizing microarray technology allows the simultaneous detection of many RNA molecules. After the binding reaction between DAZL and RNA, the RNA was purified from other inhibiting molecules through a simple spin column. Microarray analysis was used to survey the targets that flowed through the binding reaction with minimal handling of the substrate. In this way many

targets can be analyzed at the same time and the risk of losing signal through repetitive processing is avoided. An expanded list of targets would further identify pathways in which DAZ and DAZL proteins participate.

Binding affinity of protein for mRNA can be estimated using the electrophoretic mobility shift assay or EMSA [82]. The EMSA enables the researcher to determine, with some degree of precision, the binding affinity of a protein for an mRNA substrate. When differing DAZL protein constructs are used in binding studies, an instructive comparison can be made to further define the function of DAZL domains.

Materials & Methods

Growth of Cells and Purification of Proteins

NEB's ER2508 cells with the appropriate pMAL vector were grown in LB under conditions cited in Chapter III Materials and Methods. After purifying the proteins through amylose resin affinity chromatography and washing the resin with buffer, the resin with the enriched DAZL and DAZ proteins were stored at 4°C until used for binding reactions.

Binding Reaction for Microarray

Figure 35 shows a flowchart associated with preparing the RNA for use in microarrays. Four Qiagen Small Spin Columns were used for the microarray experiment. Poly A+ RNA from human testis (Clontech, described previously) was

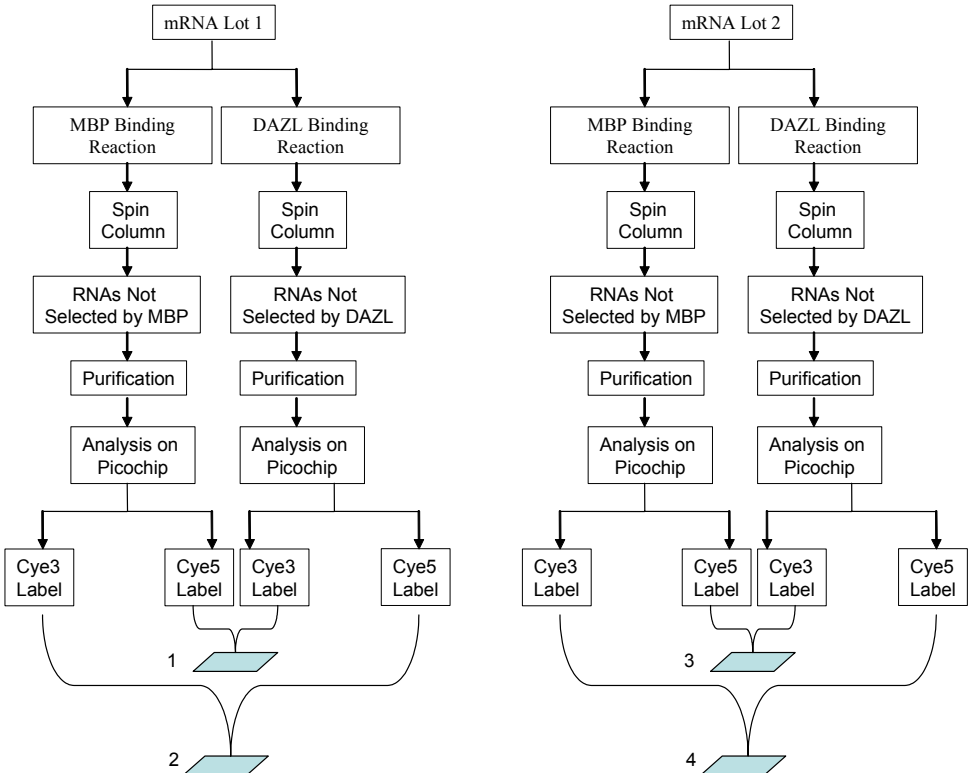


Figure 35. Flowchart of RNA Preparation for Microarray Analysis of RNAs bound by DAZL

divided into two tubes and each tube was subsequently divided into an aliquot for use with MBP-bound (maltose binding protein) resin and an aliquot for use with DAZL-bound resin. The products of one set were labeled according to manufacturer's instructions with fluorescent dyes (Amersham Biosciences, Piscataway, NJ); the dye Cy3 was used with RNAs associated with MBP and the dye Cy5 was used with RNAs associated with DAZL. The products of the second set were labeled with the alternate dyes (dye-swap).

At the Scott and White Microarray Center, Qiagen Small Spin Columns were pretreated with 1X Binding Buffer (321mM Tris, pH 7.4, 20mM KCl, 5mM $MgCl_2$,

1mM DTT, 10% Glycerol, 7.5 µg/mL yeast tRNA, 100µg/mL BSA). After pre-treatment, the tubes were centrifuged in a tabletop centrifuge at 13,000 rpm for 1 min. In separate PCR tubes, 50µl 1X Binding Buffer, pH 7.4 was added to 2.5 µg in 2.5 µl Poly A+ Human Testis RNA from Clontech (Cat# 6535-1). To this, 300 µg in 100µl of the protein-bound resin was added using a 200µl pipette tip with the end cut off to avoid resin sticking in the tip. 300µg is based on the maximum binding capacity of the amylose resin at 3 mg/ml binding of bed volume. The mixture was incubated on ice for 20 min and the complete reaction was applied to the pretreated column. The column was centrifuged at 13,000 rpm for 1 min in a tabletop microcentrifuge and the flowthrough was immediately transferred to the technicians at the Scott & White Microarray Center for downstream manipulations and analyses.

Microarray Analysis

In this study, the mRNA after centrifugation required purification due to residual proteins and other components of the binding reaction which would inhibit downstream microarray manipulations. Purification of the RNA was achieved using the Ambion (Cat# 1908, Austin, TX) Megaclear purification column. Methods followed were those provided in the manufacturer's instructions. Briefly, RNA was bound to the column, washed with buffer to remove any contaminants, and eluted in a low salt buffer suitable for many enzymatic reactions. The mRNA from each of the enriched samples was analyzed using the Agilent 2100 Bioanalyzer (Palo Alto, CA) and an RNA 6000 NanoLabChip Kit (Agilent, Cat# 5065-4474). Each of the four chips was loaded with

mRNA and analyzed using Agilent Technologies Bio Analyzer Sizing Software (Palo Alto, CA).

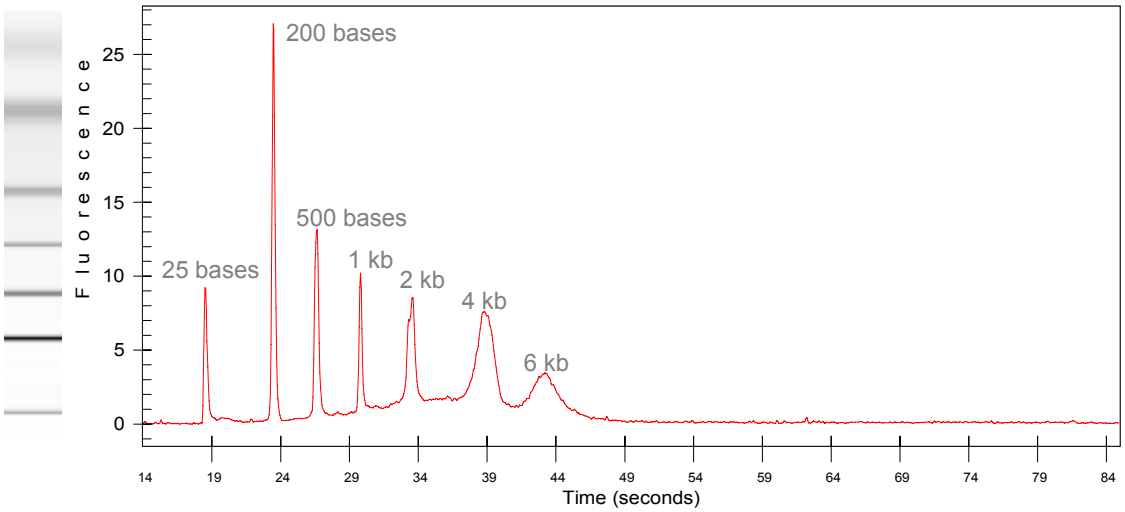


Figure 36. Capillary Electrophoresis Microarray Ladder. This graph displays, on the left, a simulated gel run of an RNA from capillary electrophoresis. On the right is a chromatogram showing the intensity of peaks versus migration in time. The longer pieces of RNA take longer to migrate.

Figure 36 shows the results of use of Agilent Technologies' capillary electrophoresis chip for determining quality of RNA. A simulated gel picture of the run is shown on the left. This picture is simulated because the actual technology relies on capillary electrophoresis and fluorescent dye which is associated with the electrophoresed sample. A laser scans the migration of the nucleic acids and the output is a chromatogram indicating size and intensity of bands. In Figure 37, the fluorescence data are on the right and are quantitative showing a ladder which is used in sizing

unknown samples. Amplification of the RNA proceeded after the RNA quantity and quality was of a sufficient nature.

The previous RNA assay determined that there was 8.273 ng/ul of RNA which flowed through MBP treatment 1, 6.853 ng/ul of RNA from MBP treatment 2, 5.505 ng/ul of RNA from DAZL treatment 1, and 5.185 ng/ul of RNA from DAZL treatment 2. Two hundred nanograms of each sample were used in the generation of labeled aRNA (amplified RNA) for the microarray using the Ambion Amino Allyl MessageAmp aRNA Kit (Cat# 1752, Austin, TX). Amplified RNA is RNA which is created from transcription from a cDNA template. The cDNA template is obtained by reverse transcription from the original RNA substrate. Reverse transcription began with an oligo (dT) primer containing a T7 promoter sequence binding to the poly A tail of the transcript. A reverse transcription enzyme, Arrayscript, synthesized the cDNA. Second strand synthesis took place next to provide a template for *in vitro* transcription which included the incorporation of the modified nucleotide 5'-(3-aminoallyl)-UTP into the aRNA. Once the aRNA (amplified RNA) was purified and reverse-transcribed, fluorescent labeling was achieved during amplification of the cDNA. For labeling, each of the two samples were split into two aliquots in which the RNAs selected using MBP resin were labeled red and the RNAs selected using DAZL resin were labeled green. In the other group, the MPB-specific RNAs were labeled green and the DAZL-specific RNAs were labeled red. The dye-swapping method is routinely used to correct for anomalies in labeling efficiency. The modified UTP could then be bound by reactive dyes such as Cy3 (green, Cat# PA25001) and Cy5 (red, Cat# 23001) (Amersham

Biosciences). After amplification, nanochip analysis was performed again for both samples and repetitions. In 100 µl, there was 0.59 µg/µl MBP 1, 0.58 µg/µl MBP 2, 0.11 µg/µl DAZL 1, and 0.16 µg/µl DAZL 2 synthesized.

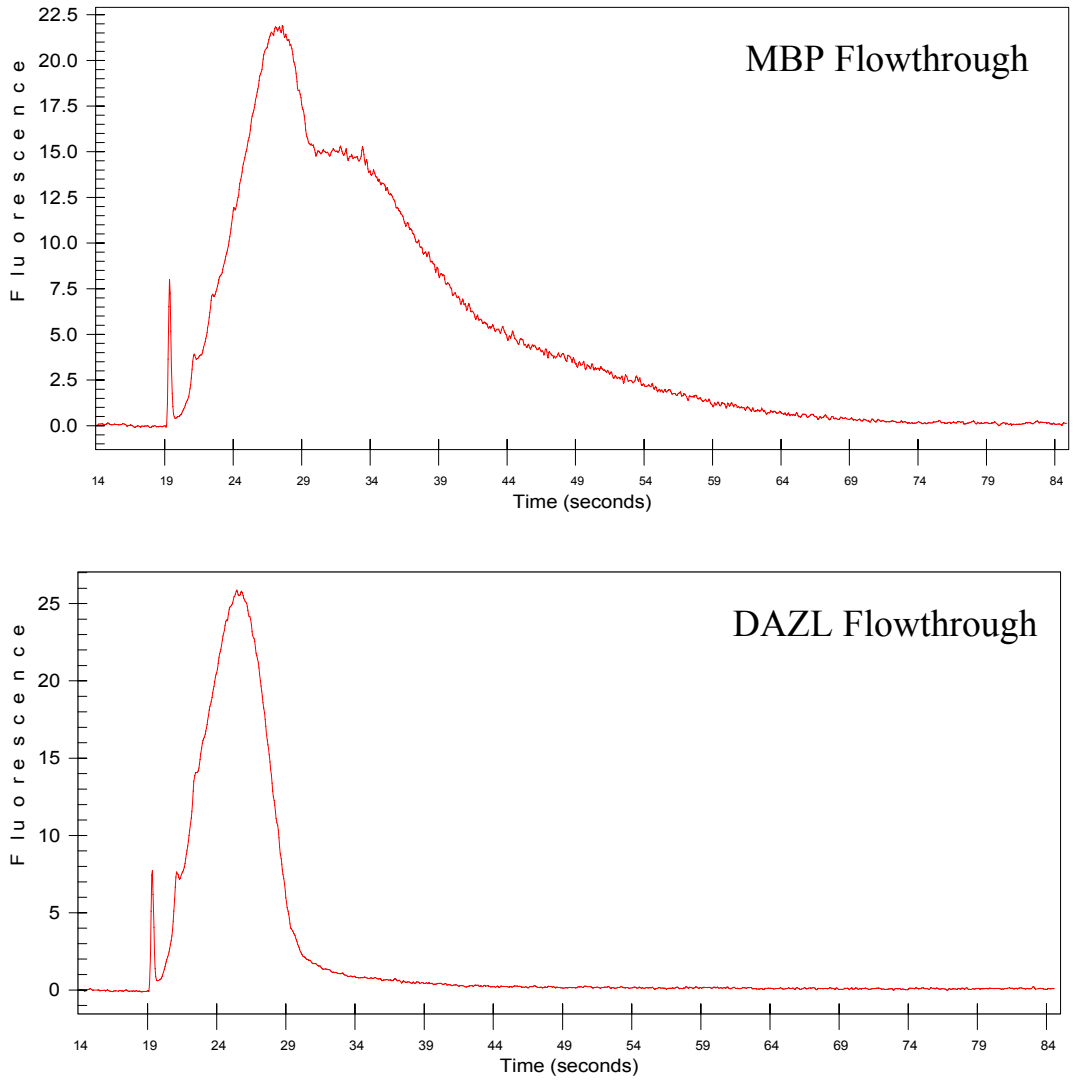


Figure 37. Lab on a Chip Analysis of Flowthrough RNA After Treatment by MBP and DAZL. The RNAs that flowed through the column after treatment with MBP showed a higher number of larger species of RNA than the column treated with DAZL. DAZL selectively retains most of the transcripts larger than 2 kb while the smaller transcript population appears to have been left intact.

For the microarray, 3 µg of each amount was applied to each of the microarrays. The mixtures of each group were as follows: 5.1 µl MBP 1 (Cy3, green) X 27.3 µl DAZL 1 (Cy5, red), 5.1 µl MBP 1 (Cy5, red) X 27.3 µl DAZL 1 (Cy3, green), 5.2 µl MBP 2 (Cy3, green) X 18.8 µl DAZL 2 (Cy5, red), and 5.2 µl MBP 2 (Cy5, red) X 18.8 µl DAZL 2 (Cy3, green). The paired samples were fragmented using components found in the kit from Ambion (Cat# 8740). The fragmented samples were hybridized overnight with hybridization buffer from Ambion (SlideHyb, Cat# 8861) at 48ºC to the Human21K Oligo Library V2.1 (Operon, Huntsville, AL). The microarray plates were read in a GenePix 4000A scanner (Axon Instruments, Sunnyvale, CA) and the data were normalized and filtered using GeneSpring v7.2 software package (Silicon Genetics, Agilent Technologies). Gene expression was assessed using multiple t-tests ($p < 0.05$) using Benjamini Hochberg false discovery rate multiple testing correction. From these data the GeneSpring software determine pathways and functions which DAZL would influence by binding RNAs.

Genes Selected and Not Selected by DAZL for in vitro *Transcription*

Genes specifically selected by the DAZL protein from the microarray assay were used as targets to create transcripts to verify the binding system. One of the mRNAs preferentially selected by DAZL was PSMD (Proteasome 26S Subunit 6); (Genbank Accession #NM_014814) and its cDNA was generated using PCR with the addition of a T7 promoter on the 5' end. Also, a gene not selectively retained in the DAZL column was Testis-Specific Protein Kinase 1; TESK1 (Genbank Accession # NM_006285). *In*

vitro transcription was used to generate a radiolabeled transcript of the PCR product for each of these mRNAs using methods previously described. Both of the probes originated from PCR from the Clontech cDNA Testis Library using the following primers: For TESK1 (low binder) Forward Primer, NM_006285 5F,- 5'- TTG GCT TCC TGT CCA TGT GC -3', TESK1 Reverse Primer, NM_006285 5R, - 5'- GCC AGC CAG GCA GTT TTA TTG, PSMD Forward Primer, NM_014814F+T7, -5'- TAA TAC GAC TCA CTA TAG GAG AAC CTG GAG GAG GAG -3', and PSMD Reverse Primer, NM_014814R, -5'- GAA ACA ACA ATG CAG TAT TTA TTT TAT ACA GCT GAC CTG GGC -3'. Each of the PCR products was cloned into the pCR2.1 vector and verified as to size and identity through sequencing. Minipreps were made of each of the clones.

For PSMD, the plasmid was linearized with ApaI (NEB, Ipswich, MA) following procedures previously described in Chapter III and PCR was performed using the same primers as described. This new PCR product was used as template to support *in vitro* transcription using Ambion's T7 Polymerase in the presence of ^{32}P-UTP.

For TESK1, PCR was performed using new primers containing a T7 promoter as the original cloning primers did not contain a T7 promoter sequence. The T7-containing forward primer, NM_006285 F+T7.2, possessed the sequence 5'- TAA TAC GAC TCA CTA TAG GTT GGC TTC CTG TCC ATG TGC -3' and the reverse primer, NM_006285 RMod.2, contained the sequence 5'- GCC AGC CAG GCA GTT TTA TTG AAA TCT TTT TAA ATA ATT GCA CGT G -3'. The PCR product was used as template in an *in vitro* transcription reaction.

Both transcripts were verified for size and purity using a 5% Urea-PAGE gel in a BioRad Mini Protean II electrophoresis chamber. Formulas and conditions used for running these gels were from instructions in the MAXIscript Kit (Catalog # 1312, Ambion, Austin, TX).

Binding Reactions for Target mRNAs

A binding reaction was created for each of the protein constructs (MBP, DAZL, DAZLmut1, DAZLmut2, and DAZ) using a method similar to that for the binding reaction performed for the microarray. The final concentrations of the buffer for the binding reactions were 20 mM Tris, pH 7.4, 50 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 10% glycerol, 7.5 µg/ml Yeast tRNA (Ambion, Austin, TX, Cat#7119), and 0.1 mg/ml BSA (Ambion, Austin, TX, Cat#2616). All binding reactions were performed in five separate PCR tubes. As in the previous binding reactions with the spin columns, tubes were treated with 100 µl of 1X binding buffer. A binding reaction consisting of 5 µl of water, 40 µl of 2X binding buffer containing tRNA and BSA, 5 µl of RNA (final concentration in reaction was 37.5 nM for TESK1 transcript and 26.2 nM for PSMD transcript), and 30 µl of resin was added to separate PCR tubes. The reactions were allowed to proceed on ice for 1 h at which time the solutions were applied to pretreated small spin columns (100 µl of 0.1 mg/ml BSA) and centrifuged for 2 min at 800 x g at room temperature. The flow-through was pipetted into scintillation vials containing 5 ml of Ecoscint XR (National Diagnostics) and assayed for radioactivity based on CPM. The binding data were corrected for moles of protein based on calculations that follow.

To account for differing amounts of proteins on the resin, 30 µl of the resin was mixed with 1ml of column buffer (20 mM Tris-HCl, pH 7.4, 200 mM NaCl, and 1 mM EDTA) containing 10mM maltose. The solutions were read in a spectrophotometer and absorbance at 280 nm used to determine concentrations of the proteins using the molar extinction coefficient of each protein. The extinction coefficient of each protein was calculated based on data reported by Pace et al. [90]. The final concentrations of proteins in the binding reaction, based on elution of proteins from the beads, were: MBP- 12.9 µM, DAZL- 15.8 µM, DAZLmut1- 16.2 µM, DAZLmut2- 23.0 µM, and DAZ- 13.6 µM.

Creation of Radiolabeled Proteins

In vivo radiolabeling of proteins began by growing cells to a midlog phase in LB. The cells were pelleted and then washed twice in M9 medium (per liter: 6.78g anhydrous disodium phosphate, 3.0g monopotassium phosphate, 0.5g sodium chloride, and 1.0g ammonium chloride). The cells were pelleted, resuspended in M9 media containing 0.01 g/100 ml of each amino acid except methionine, 0.002 g/100 ml thiamine, 0.2 g/100 ml glucose, 0.23 g/100 ml IPTG, and 0.01g/100 ml carbenicillin. The cells were exposed to 0.3 mM IPTG final to induce expression of proteins for 2 hours and then cells were centrifuged and the pellet stored at -20°C overnight. Cell pellets were diluted into column buffer (NEB- 20mM Tris, pH 7.4, 1mM EDTA, and 200mM NaCl) for enrichment and sonicated 30 times (15s on 15s off). The proteins were then purified as per NEB instructions and have been described previously. Specific activity was

determined by taking an aliquot of the radiolabeled protein and adding that sample to 5 ml of Ecoscint XR. The sample was then associated with protein concentration using the extinction coefficient as mentioned previously.

Gel Shifts

Gel shifts were performed using agarose gels utilizing a 0.5% SeaKem Gold (BioWhittaker) gel in 0.5X TBE. Low percentage agarose gels were needed because of the large size of the probes used in the binding assay for PSMD6 and TESK1. The gels were pre-run at 70V for about 30 min prior to loading and loaded when running. The gels were run at 70V for about 3 h and then treated in a solution of 12% methanol and 10% acetic acid for about 30 min with gentle shaking. Afterwards the gel was laid on Whatman paper, covered on the top side with plastic wrap, and dried on a gel drier for 1 h without heat and then 1 h with heat. After drying, the gel was exposed to a phosphorimage screen for as long as needed (usually overnight for low signal counting). The phosphorimage screen was read in a BioRad Molecular Imager Pro (Hercules, CA).

Results

To assay the targets of DAZL, human testis poly A mRNA flow through from columns containing immobilized DAZL or MBP (control) were applied to a microarray for analysis. Figure 38 shows dot plots of fold signal increase for both DAZL and MBP. Each dot above the blue line indicates greater than a 3-fold increase with a p value < 0.05. A 3-fold increase was selected as a criterion because the noise of the species

retained by MBP was present out to 2-fold. Therefore, signals greater than 3-fold and possessing a p-value < 0.05 were considered as valid targets selected by DAZL. A total of 1,313 targets were specifically selected by the DAZL protein, but not by the MBP system (Figure 39).

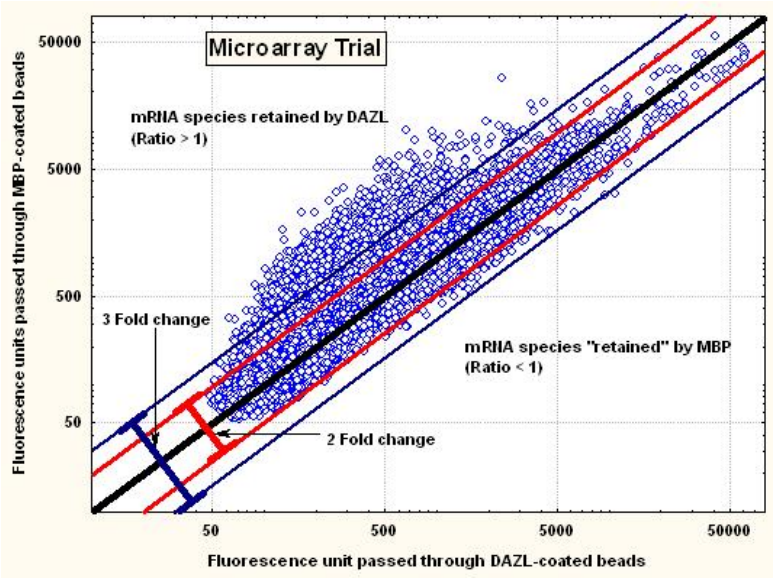


Figure 38. Species of RNA Retained by DAZL and "Retained" by MBP. The blue line indicates greater than 3-fold ratio of binding and the red line indicates greater than 2-fold ratio of binding.

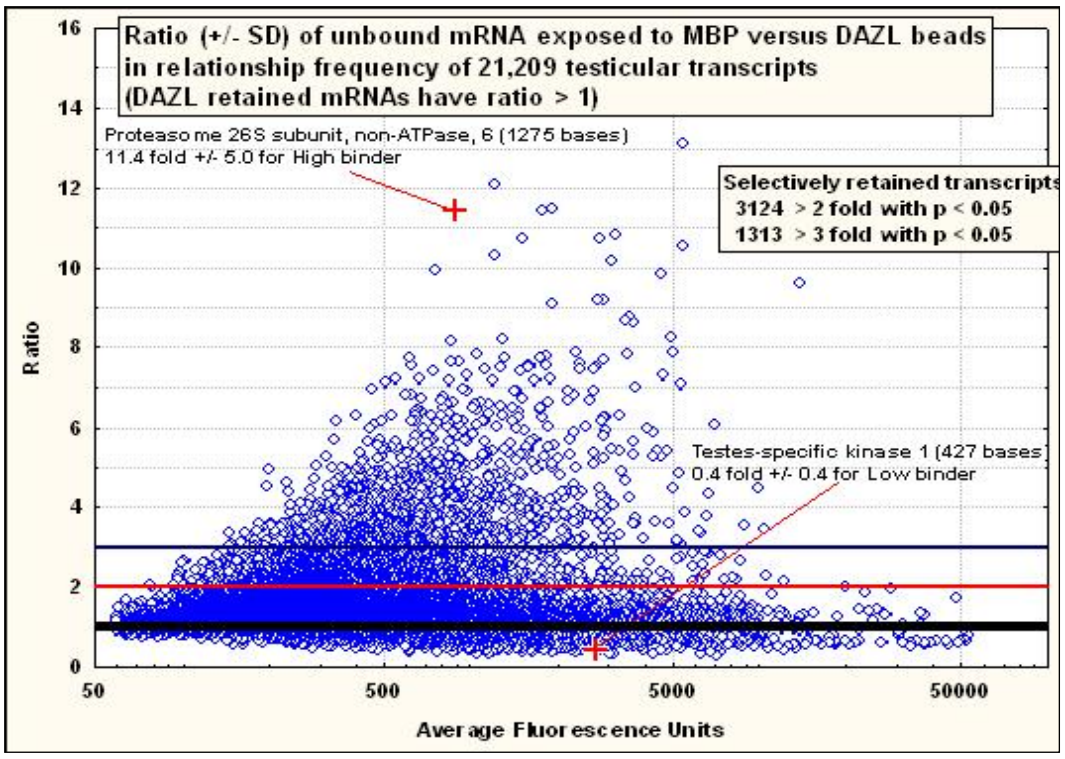


Figure 39. Possible Targets for DAZ Protein Binding. The x-axis indicateds fluorescence units and the y-axis shows the ratio of signal. One of the targets specifically selected by DAZL is PSMD6 and one of the targets not specifically bound is TESK1.

Figure 39 shows two genes selected and not selected by the DAZL protein. The top protein is the 26s proteasome subunit (PSMD), involved in degrading ubiquinated cell cycle proteins such as CDC25A. One of the least-selected transcripts by DAZL in this experiment was TESK1, a testes-specific kinase. A complete list of the top three-fold bound messenger RNAs for testicular proteins is presented in Appendix 4 and the Top 20 targets are presented in Table 11. The microarray data were compared with published results. One of the first papers to describe targets of DAZL was by Jiao et al.

who reported a number of targets of DAZL in mice using a technique called SNAAP (Specific Nucleic Acids Associated with Proteins)[88]. Table 12 shows a comparison between results of the present study and those reported by Jiao et al. Pam, Trf2, and Grsf1 were targets specifically selected by the human DAZL protein with a p-value < 0.05 in the present study and by murine DAZL using different methods by Jiao et al. This study found CDC25C to be a significant target of DAZL binding, while Jiao et al. reported minimal binding to murine DAZL. Similarly, CDC25A was reported to be bound by murine DAZL, but human DAZL showed no significant binding. Other factors not significantly selected by DAZL were; Tpx-1, associated with the process of acrosome development and sperm tail movement; Pam, a protein know to interact with Myc (a cell proliferation protein); Trf2, a factor supporting telomere maintenance; Grsf1, a positive regulator of transcription; Cappβ1, mouse actin capping protein β subunit gene which controls actin assembly; and Pα 7/C8, proteasome α 7/C8 subunit. Functions associated with these proteins could influence spermatogenesis and cell cycle control.

In addition, results were compared to those of Fox et al. in Table 13 where targets of DAZL binding in humans were studied[79]. The first group are Fox et al. selected genes considered to be significant binding targets with a low p-value in the current microarray study. The second group of genes are also those identified by Fox et al., but not statistically significant in the present study. The third group of genes was not

Table 11. Top 20 Three-Fold Targets of DAZL.

Systematic ID	Reference Sequence	Product	Biological Process	Normalized	Ratio	P-value†
NM_014814	NM_014814	proteasome regulatory particle subunit p44S10	ATP-dependent proteolysis	0.0703	14.22	0.0024
NM_000019	NM_000019	acetyl-Coenzyme A acetyltransferase 1 precursor		0.0731	13.67	0.0012
AK057433	NM_144674	hypothetical protein FLJ32871	microtubule cytoskeleton organization and biogenesis	0.0784	12.76	0.0090
NM_005977	NM_005977; NM_183043; NM_183044; NM_183045	ring finger protein 6 isoform 1; ring finger protein 6 isoform 2		0.0811	12.33	0.0011
NM_006197	NM_006197	pericentriolar material 1		0.0812	12.31	0.0017
NM_004872	NM_004872	thymic dendritic cell-derived factor 1		0.0835	11.97	0.0012
NM_014932	NM_014932	neuroligin 1	synaptogenesis; protein targeting; cell adhesion; synaptic vesicle targeting; calcium-dependent cell-cell adhesion; ion channel clustering; regulation of neuron differentiation	0.0853	11.73	0.0013
NM_007051	NM_007051; NM_131917	FAS-associated factor 1 isoform a; FAS-associated factor 1 isoform b	apoptosis	0.0870	11.50	0.0010
NM_007204	NM_007204	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	assembly of spliceosomal tri-snRNP; mRNA processing	0.0889	11.25	0.0023
AK056446	NM_005348	heat shock 90kDa protein 1, alpha		0.0971	10.30	0.0038
NM_030933	NM_030933	chromosome 1 open reading frame 14		0.0981	10.19	0.0039
NM_014633	NM_014633	SH2 domain binding protein 1	regulation of transcription, DNA-dependent	0.0982	10.18	0.0041
NM_003295	NM_003295	tumor protein, translationally-controlled 1		0.1052	9.51	0.0057
NM_001006	NM_001006; NM_182777	ribosomal protein S3a	protein biosynthesis	0.1061	9.43	0.0074
BC017296	NM_144665	sestrin 3	cell cycle arrest	0.1075	9.30	0.0016
NM_006554	NM_006554	metaxin 2	mitochondrial transport; intracellular protein transport	0.1089	9.18	0.0046
NM_003299	NM_003299	tumor rejection antigen (gp96) 1	response to stress; protein folding	0.1096	9.12	0.0031
NM_006461	NM_006461	mitotic spindle coiled-coil related protein		0.1097	9.12	0.0063
NM_002592	NM_002592; NM_182649	proliferating cell nuclear antigen	DNA repair; DNA replication; cell proliferation; cell cycle control	0.1109	9.01	0.0025
NM_053023	NM_053023; NM_170768	zinc finger protein 91 isoform 1; zinc finger protein 91 isoform 2		0.1109	9.01	0.0306

The first two columns indicate the accession numbers found in Genbank. The product column describes the gene name and the biological process column indicates which pathways are involved in the process. The normalized column contains values generated from the ratios of red to green fluorescence averaged over the duplicate runs found in the experiment. The ratio column is merely the 1/normalized ratio.

† The p-value is a statistical test as compared to a ratio of 1.0.

Table 12. Comparison of Targets of Human DAZL with Targets of Murine DAZL.

Mouse ID	Binding to mDAZL*	Human ID	Retained by hDAZL Ratio (SD)	P-level
Cdc25A	Yes	CDC25A	1.4 (0.4)	NS
Cdc25C	minimal	CDC25C	2.7 (0.9)	0.0005
Tpx-1	Yes	CRISP1	1.1 (0.4)	NS
Pam	Yes	PAM	2.8 (0.6)	0.0001
Trf2	Yes	TBPL1	4.7 (1.5)	0.004
Grsf1	Yes	GRSF1	4.3 (1.2)	0.001
Cappbeta1	Yes	CAPZB	1.5 (0.3)	NS
Palpha7/C8	Yes	PSMA3	0.9 (0.2)	NS
H47	Yes			ND
Clone D2	Yes			ND

*Jiao et al. [88]

represented in the arrays used in the current study, but were reported by Fox et al. Of note is that Fox et al. presented some targets for DAZL binding as protein sequences. A Genbank database search was used to assign a cDNA accession number for protein

Table 13. Comparison of Targets of Human DAZL with Fox, et al.

Fox	Name	Ratio	p-value
NM_183380	Bullous pemphigoid antigen 1	4.18	0.009
NM_021190	Polypyrimidine tract binding protein 2	6.87	0.010
NM_022333	TIA1 cytotoxic granule-associated RNA-binding protein-like 1	5.76	0.012
NM_004362	Calmegin (CLGN)	3.56	0.013
NP_037534, NM_013402	Fatty acid desaturase 1 (FADS1)	2.60	0.014
NM_024835	Zinc finger protein 403 (ZNF403)	5.49	0.017
NM_00210	Integrin a6 precursor (VLA-6) (CD49f)	2.55	0.018
NM_002762	Protamine P2	5.47	0.019
NM_004457	Acyl-CoA synthetase long-chain family member 3 (ACSL3)	3.29	0.020
NM_000365	Triosephosphate isomerase 1 (TPI1)	4.72	0.021
NP_115789, NM_032413	Normal mucosa of esophagus specific 1	3.11	0.022
NP_004492, NM_004501	Heterogeneous nuclear ribonucleoprotein U	2.07	0.022
NM_003103	NREBP, SON binding protein	3.08	0.026
NM_006263	Proteasome (prosome, macropain) activator subunit 1 (PA28 a)	2.24	0.031
NP_005768, NM_005777	RNA-binding motif protein 6	2.17	0.033
NM_007195	Polymerase (DNA directed) L (POLI)	3.01	0.035
NM_016614	TRAF and TNF receptor-associated protein (AD022)	2.64	0.039
NM_003054	Solute carrier family 18 (vesicular monoamine), member	1.00	NS
NP_003065, NM_003074	Uncharacterized protein with weak homology to SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily C, member 1 (SMARCC1)	1.73	NS
NM_001819	Chromogranin B (secretogranin 1)	1.45	NS
AK057058	Hypothetical protein FLJ32496	1.04	NS
NM_004147	Developmentally regulated GTP-binding protein 1 (DRG1)	1.75	NS
NM_020655	Junctophilin 3	1.66	NS
NM_000157	Glucosidase, h; acid (includes glucosylceramidase)	1.89	NS
NP_060585, NM_018115	Hypothetical protein FLJ10498	1.52	NS
NM_002761	Protamine 1	1.69	NS

Table 13. Continued.

Fox	Name	Ratio	p-value
NM_004450	Enhancer of rudimentary (Drosophila) homolog (ERH)	0.85	NS
NM_006568	Growth regulatory with ring finger domain (CGRR19)	1.47	NS
NM_005831	Nuclear domain 10 protein NDP52	1.73	NS
NM_021940	Stromal membrane-associated protein 1 (SMAP1)	1.59	NS
AK056729	Hypothetical protein FLJ32167	1.65	NS
NM_032037	Serine/threonine protein kinase (SSTK)	0.34	NS
NM_000484	Amyloid h (A4) precursor protein	1.57	NS
AK021619	Hypothetical protein FLJ11557	1.30	NS
NM_024610	HSPB (heat shock 27 kDa) associated protein 1	2.05	NS
NM_020411	G antigen, family D, 2 (GAGED2)	1.60	NS
NM_004184	Tryptophanyl-tRNA synthetase (WARS)	1.49	NS
NP_002002, NM_002011	Fibroblast growth factor receptor 4 (FGFR4)	1.44	NS
NM_001173	Rho GTPase activating protein 5 (ARHGAP5)	1.39	NS
NM_024085	Hypothetical protein FLJ22169	2.23	NS
NM_021179	Hypothetical protein LOC57821	1.95	NS
P09430	Spermatid nuclear transition protein 1 (STP-1)(TP-1)		ND
NM_006098	Guanine nucleotide binding protein (G protein), β-polypeptide 2-like 1 (RACK1: receptor of activated protein kinase C)		ND
BC035631	Solute carrier family 39 (metal ion transporter)		ND
AK092893	MLLT6: myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to 6		ND
NM_016424	Cisplatin resistance-associated overexpressed protein (LUC7A); homolog of yeast Luc7p, a splicing factor necessary for vegetative growth		ND
BC034491	Hypothetical protein MGC26598; no known domains		ND
AL049461	DKFZp586F0922		ND
BC046190	Hypothetical protein DKFZp434C0631		ND
NM_001477	G antigen 7B (GAGE7B)		ND
NM_001402	Elongation factor 1-a1		ND
AK022972	Hypothetical protein FLJ12910		ND
NM_000368	Hamartin (tuberous sclerosis 1 protein)		ND
NM_015002	F-box only protein 21 (FBXO21)		ND
NM_017455	Stromal cell-derived receptor-1 a (SDFR1)		ND
XM_086402	Hypothetical LOC149018		ND
NM_014786	Rho guanine nucleotide exchange factor (GEF) 17		ND
BC018132	Hypothetical protein FLJ22955		ND
NM_003722	Tumor protein p73-like (TP73L)		ND
XM_376178	Thyroid hormone receptor interactor 12 (TRIP12)		ND

The first column in this table indicates the accession number found in Genbank, the second column indicates the name of the gene listed, the third column indicates the ratio of signal found in the present study, and the fourth column shows the p-value of that signal in the present study in comparison to a ratio of 1.0.

sequences reported by Fox et al. and permit a side by side comparison with findings from the present study.

In addition, a complete list of genes with a 3-fold or greater ratio selected by the DAZL protein is included in the appendix. Table 11 lists the top 20 genes. This list of 1,313 genes adds to the currently known list of targets of DAZL binding in humans.

The results of the microarray were subjected to analysis by Genespring Software (Agilent Technologies, Santa Clara, CA) which organized the genes selected by DAZL based on known pathways in cellular processes. Table 14 shows four pathways in which DAZL binding likely plays some level of control: ribosome pathway, proteasome pathway, cell cycle, and pyruvate metabolism. Of interest is the p-values associated with each pathway indicating the likelihood that overlapping genes selected by DAZL are by chance.

Table 14. Pathways Associated with DAZL-Selected Transcripts.

P-value	Overlapping Genes	Pathway	# Genes in Pathway
7.79E-08	35	Ribosome - Homo sapiens	81
8.28E-06	18	Proteasome - Homo sapiens	32
0.00022	56	Cell Cycle - Homo sapiens	204
0.0121	18	Pyruvate Metabolism - Homo sapiens	48

The first column indicates the random chance that all genes selected by DAZL in this experiment was by chance. Overlapping genes are those selected by DAZL. The pathway indicates the one involving genes selected by DAZL. The number of genes in the pathway are those programmed into Genespring software.

The data were then arranged as a series of figures showing where the various DAZL-selected genes could play a role in their respective pathways (Figures 40, 41, 42, and 43). For each of figure there is a color bar indicating fold-change (y-axis), as well as p-value (x-axis). In each figure, there is a bar on the right-hand side indicating fold-change over MBP. This means that for each of the genes selected, DAZL bound a certain proportion of that gene more times than MBP. In addition, the location of each of the genes is displayed diagrammatically so that for a structural system each of the components can be shown with respect to spatial relationships to each other.

Once targets for DAZL were identified, a test was devised to assess how separate RNA targets behaved in the binding assay that generated the initial microarray data. In addition, naturally-occurring mutants of DAZL and DAZ were tested side by side in the binding assay. The PSMD gene was highly selected through the microarray by the DAZL protein whereas TESK1 had little preference for DAZL. The mRNAs for both genes were constructed and radiolabeled for use in an *in vitro* column-binding assay similar to that used in the microarray.

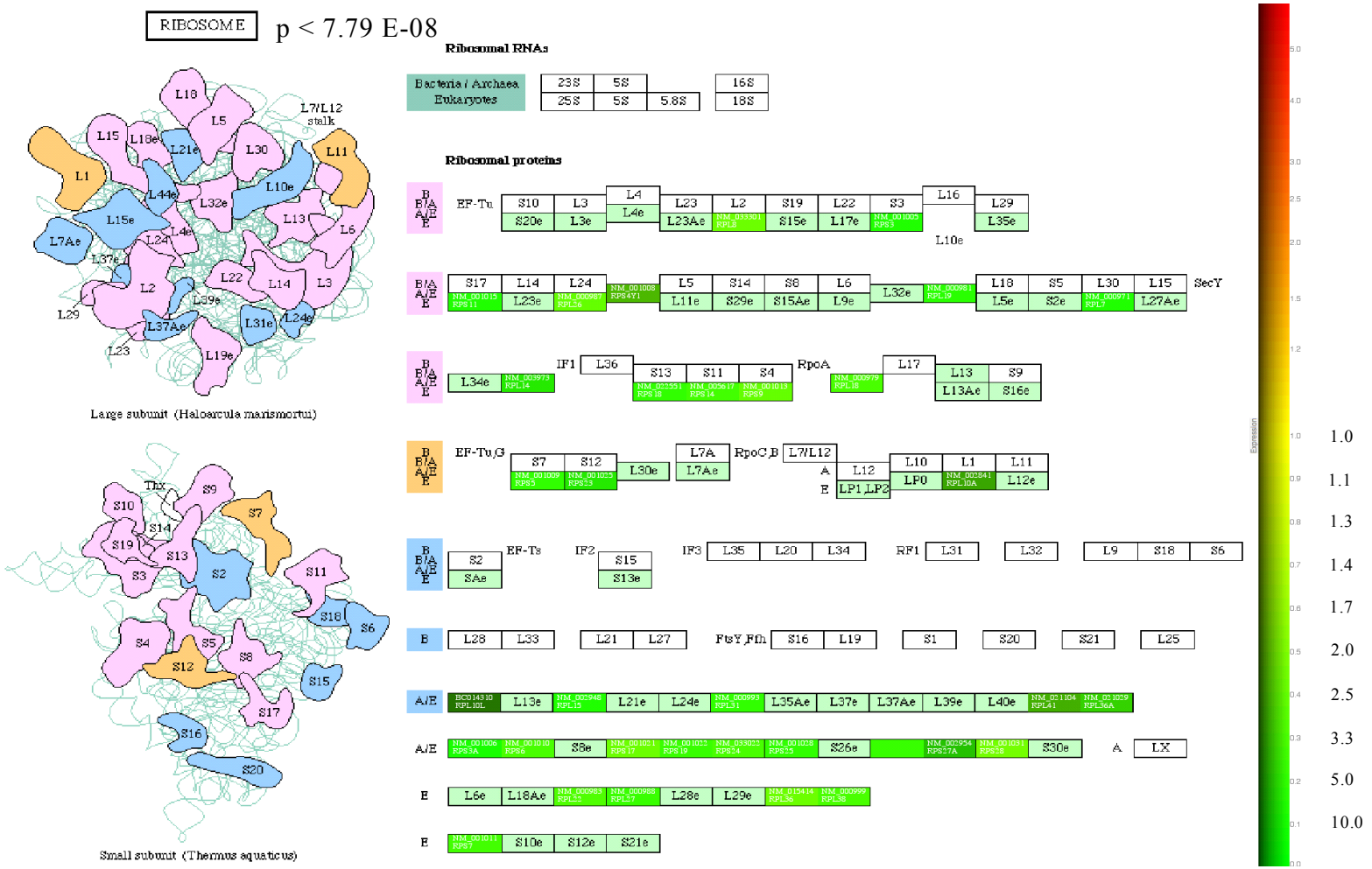


Figure 40. Ribosomal Subunits. The genes involved in ribosome construction are shown to be selected by DAZL. The deeper the shade of green, the greater the fold-change over MBP as compared with DAZL. The fold change is labeled in the green area on the right. The subunits which make up the ribosome are indicated in the middle of the figure and located within the ribosome on the left. The chance that these genes were selected at random by the experiment is shown at the top with a p < 7.79 E-08.

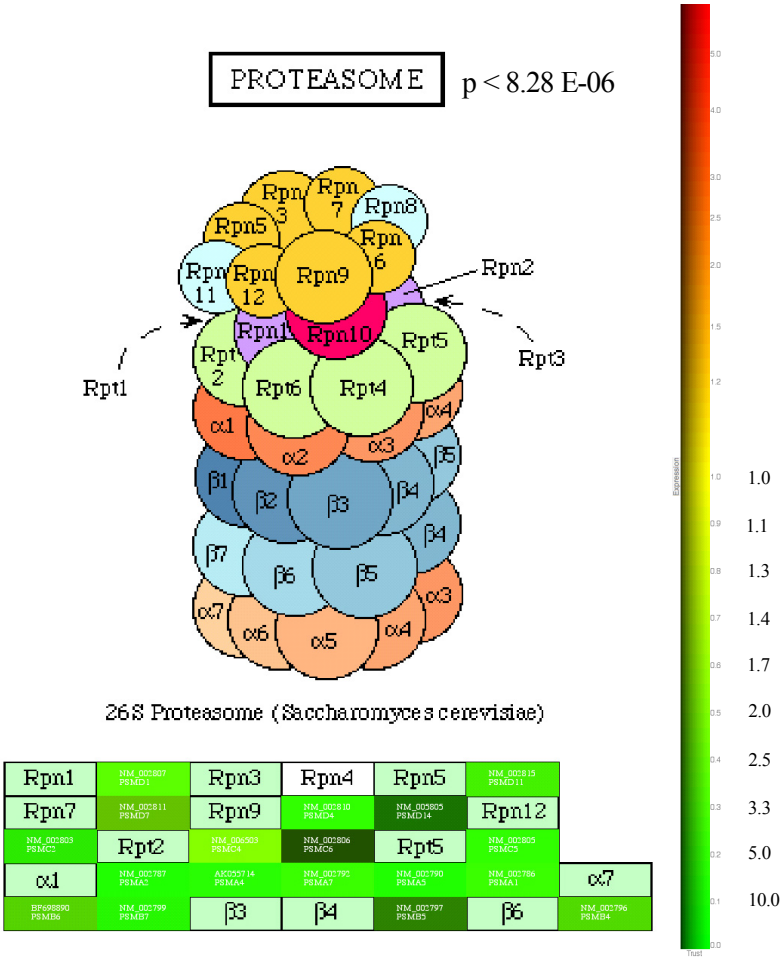


Figure 41. Genes Selected by DAZL Which Make Up the Proteasome. The fold-change over MBP is shown on the right side of the figure. The chance that these genes were randomly selected by DAZL is shown on the top with a $p < 8.28$ E-06. The components of the proteasome are indicated on the bottom with varying colors of green. The deeper the shade of green, the greater the fold-change over MBP. The locations of the subunits are indicated in the model shown above.

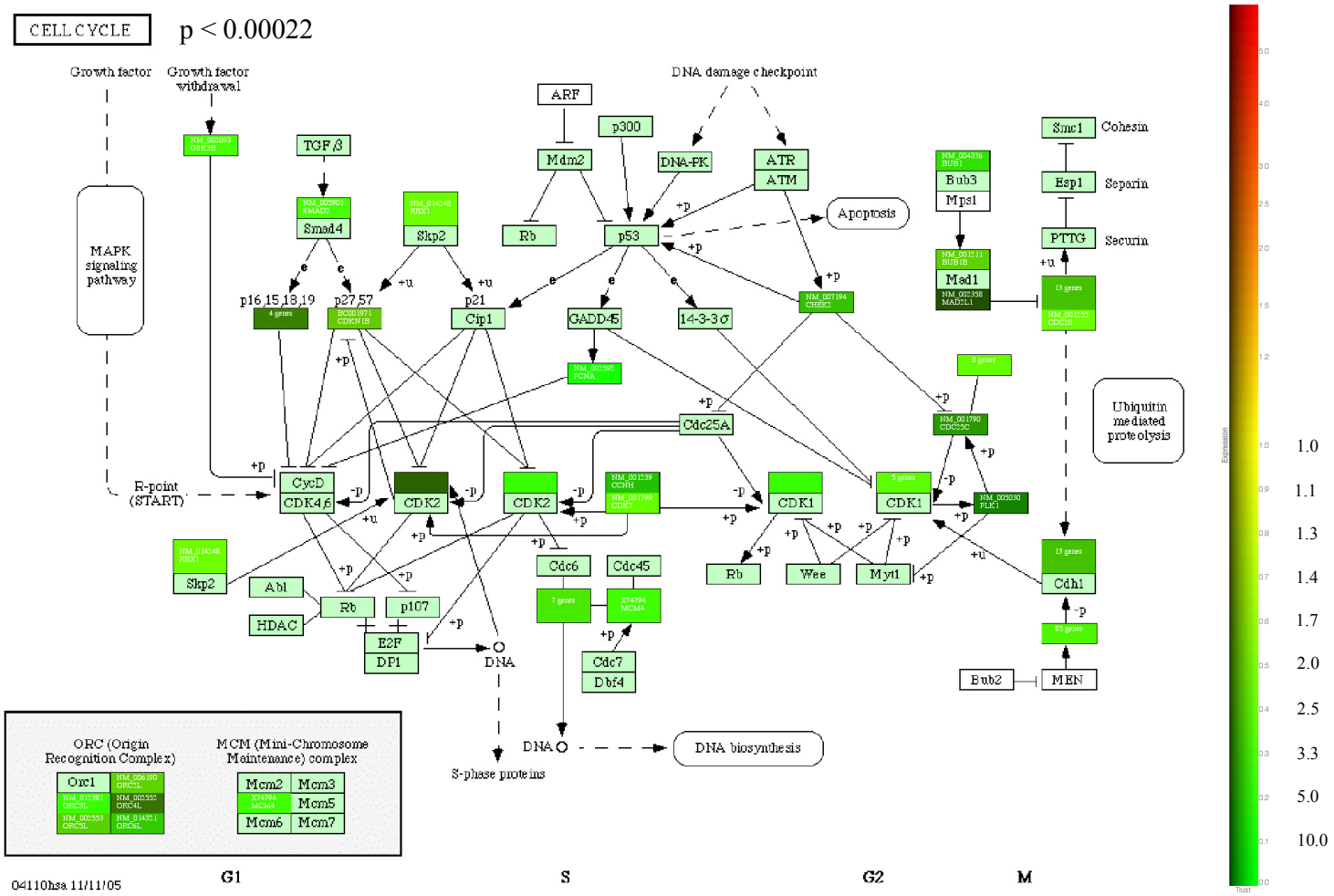


Figure 42. Genes Selected by DAZL Involved in the Cell Cycle. The fold change bar is located on the right side of the figure. The chance that these genes were randomly selected within the binding system is shown above to be $p < 0.00022$. The genes selected by the DAZL protein are shown ni the box at the bottom and the location of the genes within the displayed cell cycle pathway is indicated. The varying colors indicate the fold change over MBP where the deeper the green, the greater the fold change over MBP.

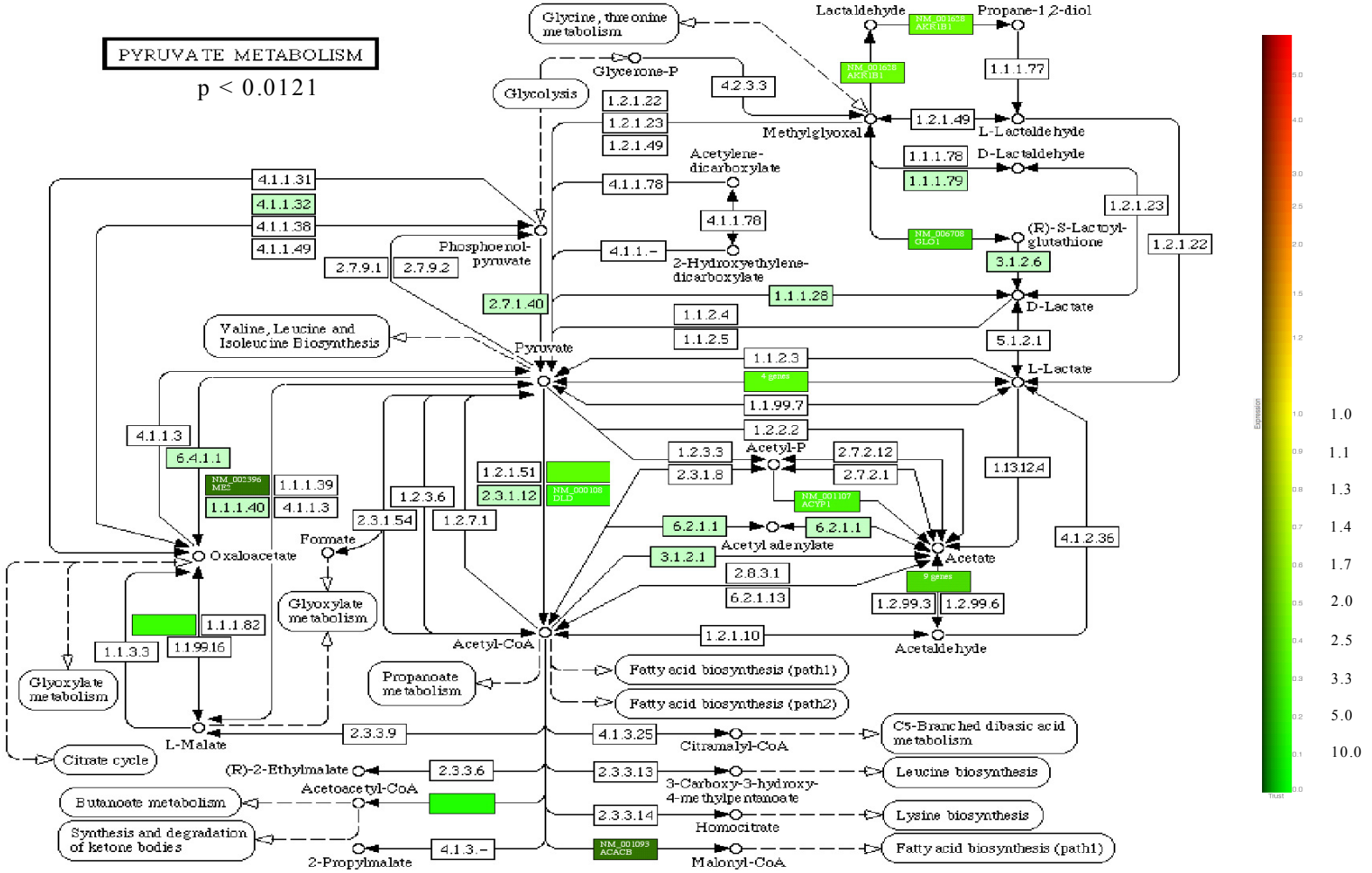


Figure 43. Genes Involved in Pyruvate Metabolism and Selected by DAZL. The bar on the right side of the figure indicates fold-change over MBP for DAZL. The deeper the shade of green, the greater the fold-change over MBP. The genes selected by DAZL and involved in pyruvate metabolism are shown in the figure. The chance that these genes were randomly selected by the DAZL binding system is shown to be $p < 0.0121$.

Table 15. Retention (p-levels) of Radiolabeled TESK1 mRNA Retained by Proteins.

Bead type	**MBP**	**DAZL**	**MUT1** Outside RRM	**MUT2** In RRM	**DAZ**
Mean counts*	399	532	310	374	468
MBP		0.009	0.086	0.526	0.095
DAZL			0.0003	0.004	0.126
MUT1				0.123	0.004
MUT2					0.065

*Radioactivity of labeled mRNA construct retained on beads after correction for moles of protein in 5 replicates for each binding reaction. The top row indicates type of protein-bound resin involved. The next row indicates counts per minute of radioactivity retained on the column. The numbers in a matrix between the bead types are p-values derived from pairwise comparisons between the different bead types using ANOVA and Dunnett's post-hoc test. The experiment is corrected for amounts of protein due to variances in binding to resin.

The column binding assay outcome represents the number of counts retained by the system (spin column + resin + proteins), calculated by subtracting the number of counts of flow-through from CPM input. TESK1, the transcript not preferred by DAZL as determined by the microarray, was first used as a probe for binding to various protein constructs. In Table 15, the second row indicates the mean counts retained on the beads after subtracting the flowthrough from the input. In this case the amount of signal retained on the resin plus the proteins for MBP is 399 CPM, whereas DAZL retained 532 CPM, DAZLmut1 retained 310 CPM, DAZLmut2 retained 374 CPM, and DAZ retained 468 CPM. The remaining rows contain a matrix of p-values which indicate pairwise comparisons between proteins. Those less than 0.05 indicate a significant

difference between the proteins. For example, the number of counts MBP retained is different ($p<0.009$) from that DAZL retains, but not from that retained by DAZLmut1 ($p = 0.086$), DAZLmut2 ($p = 0.526$), or DAZ ($p = 0.095$). When DAZL is compared with its mutants and DAZ, it becomes evident that DAZL binds more signal than DAZLmut1 ($p = 0.0003$) or DAZLmut2 ($p = 0.004$), but no more signal than DAZ ($p = 0.126$). DAZLmut1 does not bind more signal than DAZLmut2 ($p = 0.123$), but does bind less RNA than DAZ ($p = 0.004$). When DAZLmut2 is compared with DAZ there is only a trend toward a difference in binding ($p=0.065$). To summarize, MBP binds less RNA than DAZL, but MBP binding of RNA is similar to that for DAZLmut1, DAZLmut2, and DAZ. Further, DAZL binds more RNA than DAZLmut1 or DAZLmut2, but not more than DAZ. And, DAZLmut1 does not bind less RNA than DAZLmut2, but does bind less RNA than DAZ.

When the DAZL-retained radiolabeled transcript of NM_014814 (PSMD) was used as a target for DAZL another pattern of results emerged (Table 16). In this case, MBP retained 280 counts of signal, DAZL retained 678 CPM of signal, DAZLmut1 retained 515 CPM, DAZLmut2 retained 575 CPM, and DAZ retained 532 CPM. In this case MBP retained less signal than DAZL ($p = 0.0001$), DAZLmut1 ($p = 0.0002$), DAZLmut2 ($p = 0.0002$), and DAZ ($p = 0.0002$), whereas DAZL retained more signal than DAZLmut1 ($p = 0.007$), DAZLmut2 ($p = 0.031$), and DAZ ($p = 0.010$). DAZLmut1 did not bind more signal than DAZLmut2 ($p = 0.376$) or DAZ ($p = 0.696$), and DAZLmut2 did not bind more material than DAZ ($p = 0.342$). Therefore, DAZL binding is significantly greater than for any of the other protein constructs and all protein

constructs retain more RNA than MBP. Finally, DAZLmut1, DAZLmut2, and DAZ did not differ in binding RNA in this study.

To examine the Kd of DAZL and its mutant forms for the PSMD RNA probe, an electrophoretic mobility shift assay was employed. Figure 44 shows results of the gel shift using the PSMD radiolabeled probe with increasing amounts of DAZL protein, indicating a shift in amount of RNA signal with increasing amounts of DAZL protein. There is also some shift in the MBP control lane, but not in the same area as the DAZ protein shifts.

Table 16. Retention (p-levels) of PSMD mRNA Retained by Various Protein Constructs.

Bead type	**MBP**	**DAZL**	**MUT1** Outside RRM	**MUT2** In RRM	**DAZ**
Mean counts*	280	678	515	575	532
MBP		0.0001	0.0002	0.0002	0.0002
DAZL			0.007	0.031	0.010
MUT1				0.376	0.696
MUT2					0.342

*Radioactivity (CPM) retained on beads after correction for moles of protein in 5 replicates for each binding reaction. The top row indicates type of protein-bound resin involved. The next row indicates counts of radioactivity retained on the column. The numbers in a matrix between the bead types are p-values derived from pairwise comparisons between the different bead types.

Figure 45 documents the quantitative data extracted from the bands from the phosphorimage screen. DAZL under these conditions possessed a Kd of 3.1 μM of protein concentration with an R^2 of 0.82, DAZLmut1 had a Kd of 3.6 μM with an R^2 of 0.92, and DAZLmut2 had a Kd of 0.9 μM with an R^2 of 0.92. These data indicate similar Kds for each of these proteins.

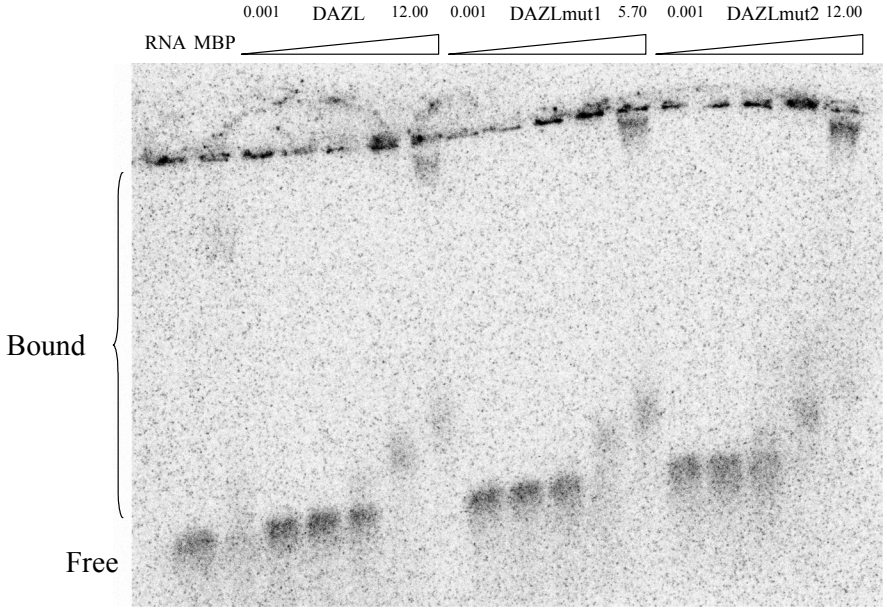


Figure 44. Gel shift Using PSMD as a Radiolabled Probe. All lanes contain 15nM final RNA concentration. The first lane represents a 20 μl reaction containing RNA only, the second lane contains 12μM MBP, the remaining lanes contain increasing amounts of protein with the lower amount indicated to the left of the label and the greatest amount indicated to the right of the label. All values are in μM.

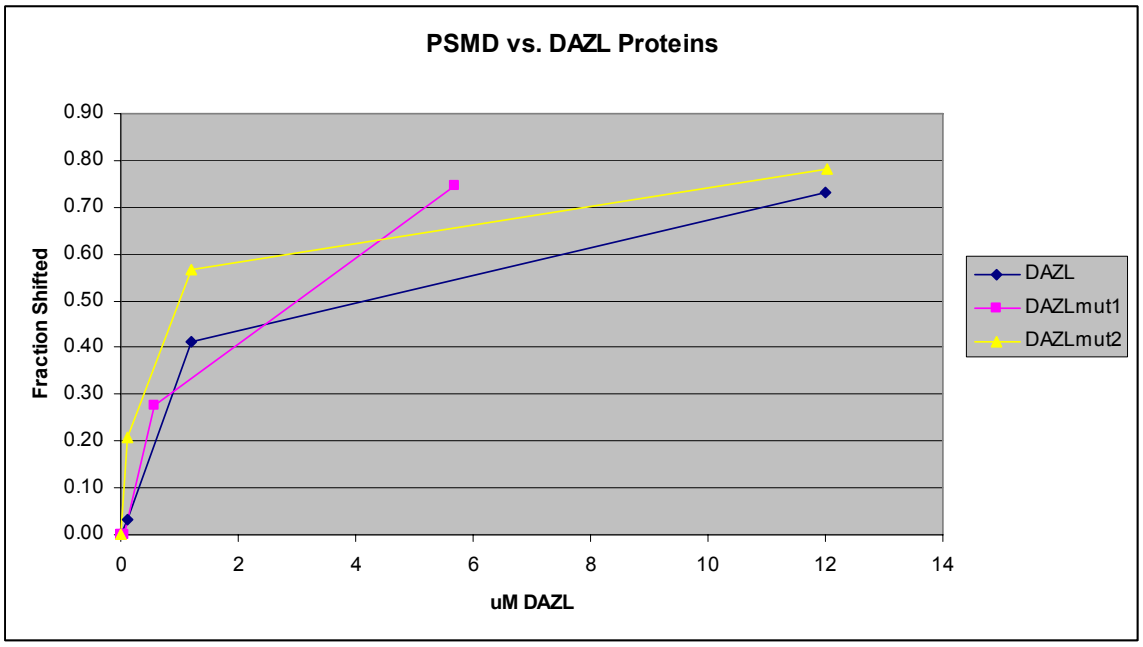


Figure 45. Gel Shift of DAZL and Mutant Forms of DAZL vs. Radiolabeled PSMD Probe. Three curves were plotted according to type of protein added to the binding reaction. A logarithmic curve was fitted to each of the series of data and the R^2 for each of the lines were DAZL = 0.82, DAZLmut1 = 0.92, and DAZLmut2 = 0.92. The Kd of each of the proteins was DAZL = 3.1, DAZLmut1 = 3.6, and DAZLmut2 = 0.9.

Discussion

This study is the first to address many aspects of function of the human DAZ family of proteins. First, microarray analyses identified a large number of genes that were specifically retained by the DAZL protein without serial enrichment of RNAs reported by Fox et al. to "enhance specificity" [79]." Their method may increase specificity for finding a consensus sequence, but may miss interactions where DAZL

proteins serve a bone fide function. In the present study, the microarray found many more targets in less time and likely at less cost. Differences between results of the current study and that of Fox et al. was not unexpected. First, the 200 colonies that they picked at random did not detect any statistically significant targets.

In 2002, Jaio et al. reported mouse targets of Dazl binding that may not directly extend to humans even though there is evolutionary conservation among species. The presence of the DAZ genes on the Y chromosome may affect the specificity of targets of Dazl binding in the mouse versus DAZL binding in humans. The mouse possesses one copy of the autosomal Dazl gene while humans have multiple DAZ genes on two different chromosomes; the human chromosome Y version is related to infertility and the role of autosomal DAZL in supporting spermatogenesis is less clear. However, a simple Genbank search reveals that the mouse Dazl gene is not 100% identical to the human DAZL gene which suggests small changes in amino acid sequence that may alter functionality. The current study revealed that MBP "binds" a relatively large number of RNAs. There are a couple of possibilities to explain this phenomenon. First, none of the proteins used in this study were highly purified to eliminate small amounts of bacterial proteins that could interact with human RNA. Because the DAZL proteins and MBP proteins were expressed in the same cell line, any "contaminating" proteins would be equivalent. Second, MBP binding may result because of the technique used to load microarrays with RNA and the fact that, after amplification, the amount of DAZL RNA was lower than the amount of MBP RNA. This had the effect of over representing genes for which MBP was selecting.

Once targets of DAZL were identified, an experiment was performed to evaluate the DAZL-preferred transcript of PSMD and the ignored TESK1 transcript. Using the high and low binders, PSMD and TESK1 RNAs respectively, and the same protocols used in the microarray, DAZL binding affinities depended on the sequence involved as a target substrate. For TESK1, the low binder, DAZL bound RNA as well as DAZ, but not as well as DAZLmut1 or DAZLmut2. This is interesting because DAZLmut1 has one mutation in the amino acid sequence of the protein, but not in the RRM. However, when the High Binder, PSMD, was used as a target sequence, DAZL bound more RNA than any other protein construct. This suggests that a mutation in the RNA-binding protein altered substrate specificity of DAZL compared to wild-type or that multiple levels of translational control were affected by these mutations.

There also may be a role for different transcripts in different species operating in slightly different ways to affect spermatogenesis. For example, spermatogenesis does not progress in the same way in mice as in humans. In mice, spermatogenesis is an ordered progression along the seminiferous tubules whereas it progresses in a mosaic pattern in humans [91]. Because of at least these differences in control of spermatogenesis between human and mouse, different pathways may control spermatogenesis including those possibly affected by DAZL RNA binding including assembly of the ribosome, the proteasome, the cell cycle, and pyruvate metabolism. If DAZL exerts some level of control on any of these pathways, the consequences would be very significant. For example, if the ribosome is affected then there would be no translation of RNAs to make proteins needed for spermatogenesis. The proteasome and

cell cycle go hand in hand in the sense that at different stages of the cell cycle, proteins need to be broken down to inhibit their activities. The cell cycle is important as spermatogenesis is the creation of more cells through a series of defined steps. Pyruvate metabolism is also an interesting area of control because it is used for energy production by developing sperm [92]. All of these pathways are vital for proper sperm production.

To explore further the interaction of DAZ proteins with RNA, the gel shift was utilized to probe the quality of binding of DAZL to RNA. The binding efficiency under the limiting conditions of the RNA suggested that about 50% of the RNA (at 15nM) binds DAZL at 3.1 μM while DAZLmut1 bound at 3.6 μM, and DAZLmut2 bound at 0.9 μM. As a comparison, at least a 1.5 nM Kd was achieved for *Xenopus* TFIIIA for 5S RNA. This low binding efficiency of DAZL for RNA may be due to several factors. First, the amount of active protein was not determined before performing this experiment, so the binding efficiency may be understated. Second, binding may be affected by the fusion tag of MBP. Third, a simple EMSA was used for generating a range for experiments involving further study into binding efficiencies [82]. Experiments into accurate Kd values involve many dilutions over a specific range and utilizing Scatchard analysis. The EMSA in this study also indicated a shift in RNA signal in the MBP lane, but MBP is not known to be an RNA-binding protein [93] and was not initially found to bind RNA. However, co purified factors did bind and shift RNA, but to a different area of the gel. To remedy this situation in future studies, the MBP tag must be excised and the binding experiments repeated for comparison, and experiments must be with DAZL protein that is not contaminated with other RNA-

binding factors. Results of binding experiments using highly purified DAZL have not been reported. The gel shift assays in the current study provide a starting point in determining the quality and quantity of binding of DAZL to RNA according to Haynes et al.[83], but more data points within the transition range are needed to compute a reliable Kd for the DAZL-RNA interaction.

Discovering targets of binding are crucial to understanding DAZ and DAZL function(s). Because there are so many targets of DAZL representing many different functions, both DAZ and DAZL are likely to influence several complex mechanisms.

It has been reported that DAZL binds its own RNA[78], which suggests that DAZL regulates its own translation via a feedback mechanism. As shown in Appendix 4, human DAZL binds its own transcript with a ratio of 7.32 ($p = 0.0064$). Interestingly, human DAZL does not bind well to the DAZ transcript in the testis library RNAs (data not shown).

CHAPTER V

CONCLUSION

This study examined how variation in genomic DNA might contribute to infertility in human males. The DAZ/DAZL family of genes was chosen as a focus because of a previously established relationship between deletions on the Y chromosome and the number of sperm produced in humans and animal models such as the mouse [63, 94-96]. While a number of investigators have used evidence for an RNA binding domain in the DNA sequence to suggest roles for proteins encoded by these genes, few studies have provided information on mRNA species that might be involved in sufficient numbers to identify pathways likely to be affected [38, 84, 97]. So binding activity and a detailed list of potential targets were not available for this family of genes so that the mechanisms for a role in spermatogenesis were unknown.

The goal of the first project was to replicate observations of others on the frequency of DAZ deletions in a population of infertile patient couples in central Texas. Based on reports of deletions of DAZ on the Y chromosome (Table 5) in males with reduced or no sperm in their ejaculates, a bank of semen residues and deletion detection assays were developed to identify a subpopulation of male candidates for genetic studies. The results indicated a lower prevalence of microdeletions than in other published reports for the population in Central Texas. The fact that the present study differed from other groups was interesting, but not entirely unexpected. One group from India demonstrated variances in deletions within the AZF region with similar testicular

etiologies within India [98] and suggested that Y chromosome fertility based on deletions in the AZF region is most likely multifactorial in origin. That is, there are multiple levels of control of genes on the Y chromosome that influence spermatogenesis in addition to Y chromosome deletions. This is likely in the case of the DAZ/DAZL family of genes as the DAZL locus is on chromosome three in humans and there is considerable variability in the range of reduced spermatogenesis with relation to complete absence of all DAZ copies on the Y chromosome.

By examining DNA at the nucleotide level, a second study was to identify SNPs (single nucleotide polymorphisms) within the DAZL gene in relation to infertility. The DAZL member of the DAZ/DAZL family of genes was selected because of its relationship to DAZ and its necessary function during spermatogenesis in mammals [22, 24, 80, 97, 99-105]. Up to this point, the pattern of study on this topic had ranged from global (whole chromosome staining from Tiepolo and Zuffardi, 1976 to detect large deleted segments of chromosomes) to point mutations in DAZL genes resulting in theoretical changes in amino acids that might alter function of the DAZL protein as in the current study [5]. In the present study, the only mutation discovered in DAZL located within an exon was a null mutation. Other mutations within introns were identified. Another group from Taiwan found mutations both within an exon and within introns of DAZL, and one was associated with reduced fertility[69]. In the current study, a SNP common within the Taiwanese population but not associated with reduced fertility, was also detected in the central Texas population. However, when the published mutations were incorporated into DAZL, the resulting recombinant protein did

have altered binding activity to RNA as demonstrated in the third project of this study. One of the mutations in question was outside the RRM of DAZL. This indicated that the complete coding sequence, rather than just the RNA-binding domain, should be examined for SNPs or other alterations including splice variants where whole exons may be deleted. Taken together, these results indicate some variation among populations with respect to genomic DNA. The larger the population examined, the more likely that additional SNPs within the DAZ and DAZL genes will be detected. This study identified reagents which were used that demonstrated the feasibility of creating recombinant proteins reflecting known SNPs and characterizing their binding properties. As binding properties change, mRNA transcription processes may be affected amplifying this single change into the action of multiple genes by altering expression.

Scanning of the intronic sequences for SNPs should be considered in a large population study, as splicing signals within DAZL or DAZ could also be affected by SNPs located within these regions. The result could potentially affect the final protein product in a manner similar to the case with a 465 C>T mutation in NADPH [72]. While the focus of this trial was to identify SNPs from individuals with reduced or absent spermatogenesis for use as the basis for a proof of concept trial, a large population study should include men with varying spermatogenesis phenotypes and scanning of the entire genomic sequence for the presence of SNPs within the four DAZ and one DAZL loci. At present, such a study may be too costly and time consuming as the introns within the DAZL and DAZ genes are quite large (~18 kbp) (Genbank NC_00003.10). As an alternative for phenotypes where some sperm cells are present in

semen, mRNAs for DAZ and DAZL can be extracted directly from ejaculated sperm cells to find splice variants. During the course of this study, but not reported here, cDNAs of DAZ and DAZL were derived from DAZL and DAZ mRNA extracted from mature, ejaculated spermatozoa [106]. This technique may allow one to search for splice variants or expressed mutations within the DAZL and DAZ genes.

To examine the function of DAZ and DAZL proteins, recombinant proteins have to be produced and used to assess mRNA-binding. Because of the inherent nature of the DAZL and DAZ proteins to bind RNA, modifications of protocols used to produce recombinant proteins were required. In addition, some cloning systems were not amenable to purification of DAZL and DAZ. Whereas other investigators have prepared recombinant proteins under denaturing conditions and renatured the proteins before examining activity, proteins used in the present study were purified under native conditions using maltose-binding protein (MBP) as a purification and solubilization tag [79]. The main problems with methods which rely on denaturation and renaturation is that there remains uncertainty as to whether the protein has folded properly at the end of the purification process and there may be substantial effort required to perfect the ideal conditions under which the protein can be purified [107]. This study utilized the pMAL system because of the simplicity in purification of native proteins. Enrichment of proteins involves the binding of the fusion protein to amylose resin and washing away the non-specific proteins. Purification utilizing the MBP fusion protein was not without its hindrances though. In this portion of the study, small quantities of bacterially-produced proteins were co-purified with the recombinant protein preparations. The

control for this process was the production of a MBP fusion protein without the DAZ/DAZL component. This important control did, however, have low levels of RNA binding which was unexpected based on the use of commercially-prepared MBP.

While several investigators have reported cloning expressed or enriched DAZL proteins from several species, there is no published literature on functional or quantitative binding of any DAZ proteins or DAZL or DAZ proteins containing amino acid substitutions [38, 78, 84, 88, 97]. Because, in at least one population, a specific mutation has been correlated with reduced spermatogenesis, this mutation provided a valuable resource for exploring DAZ or DAZL protein and RNA interactions[69]. The mutations were incorporated into proteins (labeled as DAZLmut1 and DAZLmut2) that were used for binding assays and for detecting differences in RNA binding between the various protein constructs of DAZL, DAZLmut1, DAZLmut2, and DAZ. Based on the results of these experiments, SNPs in a general population may or may not alter binding to RNA. For each variant of the DAZ or DAZL protein, the potential to impact spermatogenesis by affecting translation of different subsets of mRNAs could be examined to comparing the profile of mRNA species bound relative to the wild type.

A portion of this study was devoted to addressing the question of what testicular species of mRNA the DAZ/DAZL family of proteins bind with the expectation that the answer would suggest mechanisms for control of number of sperm cells produced. As DAZL is the only member of this family uniformly found in all mammals (DAZ is limited to old world monkeys and apes), it was used for these trials. DAZL was found to bind a substantially greater number of targets than previously reported for either murine

or human constructs [79, 88]. Others have used RNA-enrichment methods to detect a consensus sequence for DAZL binding which also identified targets[84]. It should be noted that some DAZ proteins are only found at certain stages of spermatogenesis (Figure 4), so the targets should include stages where the particular DAZ protein is present. Research into the targets of DAZ and DAZL have utilized the mouse model [84] or screened libraries in yeast to identify targets [17]. The results of those studies are not in complete agreement with those from this project as to targets for binding of DAZ or DAZL binding. Therefore, the proposed target sequences within mRNAs can not be readily deduced. Even when enrichment methods or three-hybrid screening methods are used, there is no real consensus present other than to U-rich sequences. Based on the limited number of previously reported mRNA species binding to DAZL proteins for either human or murine constructs, no obvious cellular processes or pathways are implicated. However using the novel method developed in this project, a much longer list of targets for DAZL binding were identified (Appendix 4) and four cellular processes or pathways became evident including proteins of ribosome, proteasome, cell cycle control, and pyruvate metabolism. These processes, therefore, may be enhanced or altered by DAZL binding to transcripts of spermatogonia with shortened poly-A tails. Collier et al. speculated that increased production of proteins from the transcription process results from circularizing the mRNA to increase the efficiency of ribosomal processing [38]. Increasing the number of ribosomes and proteasomes would prepare a spermatogonial cell for a series of cell divisions and drastic changes in morphology from a round cell to a spermatid. Increasing certain proteins of the cell-cycle pathway may

trigger either additional mitotic divisions prior to entry into meiosis, meiosis or both. Finally, increasing proteins for pyruvate metabolism may prepare the spermatogonial cells for use of a metabolic fuel commonly available for spermatids. DAZL also binds to DAZL mRNA transcripts so that initiation of DAZL synthesis may trigger the cascade of processes or alternatively block DAZL mRNA transcription. While further investigation of the role of DAZL binding on transcription was considered outside of the scope of this study, it is clearly an important future step for pursuing the mechanisms of DAZ/DAZL family action.

The binding of mRNA by DAZ and DAZL is most likely more complicated than simply identifying a consensus sequence. The tools created and used in this study can be used to answer further questions. What do splice variants of DAZ and DAZL bind? Do variations in affinity between DAZ/DAZL proteins for various mRNAs exist? Also, other proteins binding to DAZ/DAZL proteins may alter binding affinities or specificities and this has not been described. What subset of mRNAs do the mutations in DAZL or DAZ bind? Comparisons between these proteins should provide information on the potential for genetic variation to impact male fertility. It may also suggest control points that could be targets of spermatogonia-specific treatments to impair fertility for use as a contraceptive. DAZ and DAZL are limited in expression to spermatogonia and oogonia in fetuses and to spermatogonia and sperm at various stages of spermatogenesis in post-pubertal males [1, 75, 108, 109].

Based on the general concept that DAZ and DAZL have a role in controlling gene expression by binding transcripts with short poly-A tails to circularize and increase efficiency of ribosomal production of proteins, results of this study suggest that four pathways will be enhanced [38]. The proteasome and cell cycle are interdependent and DAZL may bind transcripts involved in each of these pathways. The cell cycle relies on the proteasome to remove unwanted factors so that the cycle progresses normally and pyruvate metabolism is needed for energy production in developing sperm cells. If DAZL is facilitating the translation of mRNAs involved in protein production, then the overall effect of DAZL may be to increase translation of all mRNAs within developing sperm cells. When a mutation forms in DAZL then the global effect would be reduced sperm production due to reduced efficiency at translating these transcripts. The role or requirement of a second set of DAZ genes to produce additional proteins only in old world monkeys and apes, has not been reconciled with any specific difference in spermatogenesis between these species and other mammals that only have DAZL. Direct comparisons of human wild type recombinant DAZL and DAZ did demonstrate differences in binding activity to certain transcripts in this project. The DAZ gene is also known in humans to exist as four copies, each with some specific differences. While variation in the number of copies of DAZ presently has not been related to failed or reduced spermatogenesis, it is conceivable that these genes when or if expressed fully obtain their mRNA binding function. While beyond the scope of this project, recombinant forms for all of the various DAZ proteins could be produced and compared using the bead-binding assay methods to screen mRNA retained relative to DAZL.

DAZL would seem to be a prototype molecule as it has been relatively conserved and has similar expression profiles across species[1].

While the results of this project are promising, there are limitations to this present body of work. Investigations into the analysis of SNPs began with the intent on identifying SNPs within the RRM of DAZL in selected patients with reduced sperm numbers. However, it became evident that one must examine exons of the RRM, as well as the whole gene. Research from Taiwan indicated that there were other regions within the DAZL gene which were important [69]. Additionally, to increase knowledge about possible mutations within the DAZL gene, the study of SNPs needs to be expanded to include a large multi-ethnic population with adequate phenotypic variations in spermatogenesis and not just fertility. Also, the entire genomic sequence should be searched for the presence of SNPs within DAZL. With more data, haplotype analyses can associate genotype with phenotype and help predict alteration of splicing function. Extraction of mRNA from mature ejaculated spermatozoa will help to confirm the presence or absence of splice variants of DAZL.

Another limitation of the results of the present study was exposed when MBP purchased from New England Biolabs failed to shift RNA while MBP enriched for the current study did shift a small amount of RNA, albeit in a non-specific manner. Because MBP was co-purified in both the MBP control and DAZL protein preps, it was decided to analyze the DAZL proteins in the various binding assays as the goal for assessing DAZL binding was to identify targets of DAZL and the co-purified contaminants present in both samples was not considered a hindrance to interpretation of results. However, it

became evident that proper evaluation of binding affinities for the various recombinant DAZL and DAZ mutants would require additional purification and separation from MBP. The constructs developed in this project contain a TEV protease site to enable the additional isolation needed for separation of recombinant proteins without using denaturing conditions.

A third concern with the project involved the size of the molecules used for electrophoretic mobility gel shift assays after microarray analysis. Both of the mRNA molecules, PSMD and TESK1, used for this experiment were full length and the consensus binding sequences within these remain unknown. This resulted in mRNAs that were to large for conventional gel shift assays and retention of material in the wells that compromised interpretation of results.

To conclude, results of this study demonstrated a reduced frequency of DAZ deletions in the general population of male partners of infertile couples in central Texas. However, genetic variation was demonstrated as point mutations in DAZL. While these mutations were not specifically associated with failed spermatogenesis, others have reported at least one such SNP. Using modifications to a variety of cloning and expression techniques recombinant proteins for DAZL, one variant of DAZ, and two DAZL mutant forms were predicted and compared. These different forms of the DAZ/DAZL family of proteins differed in binding properties. Finally, a larger number of testicular mRNA species than predicted or previously identified was found to bind to both human and wild type versions of DAZL. These mRNAs were clustered in pathways of four discrete cellular processes involved in conversion of mitotically active

spermatogonia cells to haploid morphologically specialized spermatids. These cell types are consistent with those in which DAZL and DAZ expression has been reported. These results suggest a point in the control of spermatogenesis where expression of a single molecule, DAZL, may influence a cascade of functions, and where genetic variation and perhaps future therapeutic intervention may be targeted to specifically arrest or limit spermatogenesis.

REFERENCES

1. Reynolds N, Cooke HJ. Role of the DAZ genes in male fertility. Reprod Biomed Online 2005; 10: 72-80.

2. Vogt PH. Human Y chromosome deletions in Yq11 and male fertility. Advances in Experimental Medicine & Biology 1997; 424: 17-30.

3. World Health Organization. WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. Cambridge, UK: Cambridge University Press; 1999.

4. Vogt PH. Molecular genetics of human male infertility: from genes to new therapeutic perspectives. Curr Pharm Des 2004; 10: 471-500.

5. Tiepolo L, Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. Human Genetics 1976; 34: 119-124.

6. Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC. The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science 1992; 258: 52-59.

7. Ma K, Inglis JD, Sharkey A, Bickmore WA, Hill RE, Prosser EJ, Speed RM, Thomson EJ, Jobling M, Taylor K, et al. A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. Cell 1993; 75: 1287-1295.

8. Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Grone HJ, Jung A, Engel W, Haidl G. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Human Molecular Genetics 1996; 5: 933-943.

9. Ferlin A, Tessari A, Ganz F, Marchina E, Barlati S, Garolla A, Engl B, Foresta C. Association of partial AZFc region deletions with spermatogenic impairment and male infertility. J Med Genet 2005; 42: 209-213.

10. Hucklenbroich K, Gromoll J, Heinrich M, Hohoff C, Nieschlag E, Simoni M. Partial deletions in the AZFc region of the Y chromosome occur in men with impaired as well as normal spermatogenesis. Hum Reprod 2005; 20: 191-197.

11. Yang Y, Xiao CY, A ZC, Zhang SZ, Li X, Zhang SX. DAZ1/DAZ2 cluster deletion mediated by gr/gr recombination per se may not be sufficient for spermatogenesis impairment: a study of Chinese normozoospermic men. Asian J Androl 2006; 8: 183-187.

12. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nature Genetics 1995; 10: 383-393.

13. Saxena R, de Vries JW, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JM, Page DC. Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. Genomics 2000; 67: 256-267.

14. Lin YW, Thi DA, Kuo PL, Hsu CC, Huang BD, Yu YH, Vogt PH, Krause W, Ferlin A, Foresta C, Bienvenu T, Schempp W, Yen PH. Polymorphisms associated with the DAZ genes on the human Y chromosome. Genomics 2005; 86: 431-438.

15. Teng YN, Lin YM, Lin YH, Tsao SY, Hsu CC, Lin SJ, Tsai WC, Kuo PL. Association of a single-nucleotide polymorphism of the deleted-in-azoospermia-like gene with susceptibility to spermatogenic failure. J Clin Endocrinol Metab 2002; 87: 5258-5264.

16. Yen PH, Chai NN, Salido EC. The human DAZ genes, a putative male infertility factor on the Y chromosome, are highly polymorphic in the DAZ repeat regions. Mammalian Genome 1997; 8: 756-759.

17. Tsui S, Dai T, Roettger S, Schempp W, Salido EC, Yen PH. Identification of two novel proteins that interact with germ-cell- specific RNA-binding proteins DAZ and DAZL1. Genomics 2000; 65: 266-273.

18. Draper DE. Themes in RNA-protein recognition. J Mol Biol 1999; 293: 255-270.

19. Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem 1993; 62: 289-321.

20. Nagai K, Oubridge C, Ito N, Avis J, Evans P. The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. Trends Biochem Sci 1995; 20: 235-240.

21. Ruggiu M, Cooke HJ. In vivo and in vitro analysis of homodimerisation activity of the mouse Dazl1 protein. Gene 2000; 252: 119-126.

22. Xu EY, Moore FL, Pera RA. A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans. Proc Natl Acad Sci U S A 2001; 98: 7414-7419.

23. Lin YM, Chen CW, Sun HS, Tsai SJ, Lin JS, Kuo PL. Presence of DAZL transcript and protein in mature human spermatozoa. Fertil Steril 2002; 77: 626-629.

24. Ruggiu M, Speed R, Taggart M, McKay SJ, Kilanowski F, Saunders P, Dorin J, Cooke HJ. The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature 1997; 389: 73-77.

25. Eberhart CG, Maines JZ, Wasserman SA. Meiotic cell cycle requirement for a fly homologue of human Deleted in Azoospermia [see comments]. Nature 1996; 381: 783-785.

26. Maines JZ, Wasserman SA. Post-transcriptional regulation of the meiotic Cdc25 protein Twine by the Dazl orthologue Boule. Nat Cell Biol 1999; 1: 171-174.

27. Lee K, Fajardo MA, Braun RE. A testis cytoplasmic RNA-binding protein that has the properties of a translational repressor. Molecular & Cellular Biology 1996; 16: 3023-3034.

28. Kleene KC. Patterns of translational regulation in the mammalian testis. Molecular Reproduction & Development 1996; 43: 268-281.

29. Sommerville J, Ladomery M. Transcription and masking of mRNA in germ cells: involvement of Y-box proteins. Chromosoma 1996; 104: 469-478.

30. Yen PH. A long-range restriction map of deletion interval 6 of the human Y chromosome: a region frequently deleted in azoospermic males. Genomics 1998; 54: 5-12.

31. Tse JY, Wong EY, Cheung AN, O WS, Tam PC, Yeung WS. Specific expression of VCY2 in human male germ cells and its involvement in the pathogenesis of male infertility. Biol Reprod 2003; 69: 746-751.

32. Kleiman SE, Yogev L, Hauser R, Botchan A, Bar-Shira Maymon B, Schreiber L, Paz G, Yavetz H. Members of the CDY family have different expression patterns: CDY1 transcripts have the best correlation with complete spermatogenesis. Hum Genet 2003; 113: 486-492.

33. Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 2001; 29: 279-286.

34. Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. Nat Genet 1997; 17: 154-163.

35. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC. A physical map of the human Y chromosome. Nature 2001; 409: 943-945.

36. Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 2002; 71: 906-922.

37. Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S. A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 2004; 83: 1046-1052.

38. Collier B, Gorgoni B, Loveridge C, Cooke HJ, Gray NK. The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. Embo J 2005; 24: 2656-2666.

39. Kent-First M, Muallem A, Shultz J, Pryor J, Roberts K, Nolten W, Meisner L, Chandley A, Gouchy G, Jorgensen L, Havighurst T, Grosch J. Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. Mol Reprod Dev 1999; 53: 27-41.

40. Simoni M, Gromoll J, Dworniczak B, Rolf C, Abshagen K, Kamischke A, Carani C, Meschede D, Behre HM, Horst J, Nieschlag E. Screening for deletions of the Y chromosome involving the DAZ (Deleted in AZoospermia) gene in azoospermia and severe oligozoospermia. Fertility & Sterility 1997; 67: 542-547.

41. Reijo R, Alagappan RK, Patrizio P, Page DC. Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome [see comments]. Lancet 1996; 347: 1290-1293.

42. Najmabadi H, Huang V, Yen P, Subbarao MN, Bhasin D, Banaag L, Naseeruddin S, de Kretser DM, Baker HW, McLachlan RI. Substantial prevalence of microdeletions of the Y-chromosome in infertile men with idiopathic azoospermia and oligozoospermia detected using a sequence-tagged site-based mapping strategy. Journal of Clinical Endocrinology & Metabolism 1996; 81: 1347-1352.

43. Qureshi SJ, Ross AR, Ma K, Cooke HJ, Intyre MA, Chandley AC, Hargreave TB. Polymerase chain reaction screening for Y chromosome microdeletions: a first step towards the diagnosis of genetically-determined spermatogenic failure in men. Molecular Human Reproduction 1996; 2: 775-779.

44. Stuppia L, Gatta V, Fogh I, Gaspari AR, Morizio E, Mingarelli R, Di Santo M, Pizzuti A, Calabrese G, Palka G. Genomic organization, physical mapping, and involvement in Yq microdeletions of the VCY2 (BPY 2) gene. Genomics 2001; 72: 153-157.

45. Pryor JL, Kent-First M, Muallem A, Van Bergen AH, Nolten WE, Meisner L, Roberts KP. Microdeletions in the Y chromosome of infertile men [see comments]. New England Journal of Medicine 1997; 336: 534-539.

46. van der Ven K, Montag M, Peschka B, Leygraaf J, Schwanitz G, Haidl G, Krebs D, van der Ven H. Combined cytogenetic and Y chromosome microdeletion screening in males undergoing intracytoplasmic sperm injection. Mol Hum Reprod 1997; 3: 699-704.

47. Girardi SK, Mielnik A, Schlegel PN. Submicroscopic deletions in the Y chromosome of infertile men. Human Reproduction 1997; 12: 1635-1641.

48. Kremer JA, Tuerlings JH, Borm G, Hoefsloot LH, Meuleman EJ, Braat DD, Brunner HG, Merkus HM. Does intracytoplasmic sperm injection lead to a rise in the frequency of microdeletions in the AZFc region of the Y chromosome in future generations? Human Reproduction 1998; 13: 2808-2811.

49. Foresta C, Ferlin A, Garolla A, Rossato M, Barbaux S, De Bortoli A. Y-chromosome deletions in idiopathic severe testiculopathies. Journal of Clinical Endocrinology & Metabolism 1997; 82: 1075-1080.

50. Vereb M, Agulnik AI, Houston JT, Lipschultz LI, Lamb DJ, Bishop CE. Absence of DAZ gene mutations in cases of non-obstructed azoospermia. Mol Hum Reprod 1997; 3: 55-59.

51. Liow SL, Ghadessy FJ, Ng SC, Yong EL. Y chromosome microdeletions, in azoospermic or near-azoospermic subjects, are located in the AZFc (DAZ) subregion. Mol Hum Reprod 1998; 4: 763-768.

52. Oliva R, Margarit E, Ballesca JL, Carrio A, Sanchez A, Mila M, Jimenez L, Alvarez-Vijande JR, Ballesta F. Prevalence of Y chromosome microdeletions in oligospermic and azoospermic candidates for intracytoplasmic sperm injection. Fertil Steril 1998; 70: 506-510.

53. Krausz C, Bussani-Mastellone C, Granchi S, McElreavey K, Scarselli G, Forti G. Screening for microdeletions of Y chromosome genes in patients undergoing intracytoplasmic sperm injection. Hum Reprod 1999; 14: 1717-1721.

54. Krausz C, Quintana-Murci L, Barbaux S, Siffroi JP, Rouba H, Delafontaine D, Souleyreau-Therville N, Arvis G, Antoine JM, Erdei E, Taar JP, Tar A, Jeandidier E, Plessis G, Bourgeron T, Dadoune JP, Fellous M, McElreavey K. A high frequency of Y chromosome deletions in males with nonidiopathic infertility. J Clin Endocrinol Metab 1999; 84: 3606-3612.

55. Ferlin A, Moro E, Garolla A, Foresta C. Human male infertility and Y chromosome deletions: role of the AZF- candidate genes DAZ, RBM and DFFRY. Hum Reprod 1999; 14: 1710-1716.

56. Seifer I, Amat S, Delgado-Viscogliosi P, Boucher D, Bignon YJ. Screening for microdeletions on the long arm of chromosome Y in 53 infertile men. Int J Androl 1999; 22: 148-154.

57. Kim SW, Kim KD, Paick JS. Microdeletions within the azoospermia factor subregions of the Y chromosome in patients with idiopathic azoospermia. Fertil Steril 1999; 72: 349-353.

58. Kerr NJ, Zhang J, Sin FY, Benny P, Sin IL. Frequency of microdeletions in the azoospermia factor region of the Y-chromosome of New Zealand men. N Z Med J 2000; 113: 468-470.

59. Martinez MC, Bernabe MJ, Gomez E, Ballesteros A, Landeras J, Glover G, Gil-Salom M, Remohi J, Pellicer A. Screening for AZF deletion in a large series of severely impaired spermatogenesis patients. J Androl 2000; 21: 651-655.

60. Kato H, Komori S, Nakata Y, Sakata K, Kanazawa R, Handa M, Kobayashi S, Koyama K, Isojima S. Screening for deletions in interval D16-22 of the Y chromosome in azoospermic and oligozoospermic Japanese men. J Hum Genet 2001; 46: 110-114.

61. Fujisawa M, Shirakawa T, Kanzaki M, Okada H, Arakawa S, Kamidono S. Y-chromosome microdeletion and phenotype in cytogenetically normal men with idiopathic azoospermia. Fertil Steril 2001; 76: 491-495.

62. de Vries JW, Repping S, Oates R, Carson R, Leschot NJ, van der Veen F. Absence of deleted in azoospermia (DAZ) genes in spermatozoa of infertile men with somatic DAZ deletions. Fertil Steril 2001; 75: 476-479.

63. Fernandes S, Huellen K, Goncalves J, Dukal H, Zeisler J, Rajpert De Meyts E, Skakkebaek NE, Habermann B, Krause W, Sousa M, Barros A, Vogt PH. High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 2002; 8: 286-298.

64. Ambasudhan R, Singh K, Agarwal JK, Singh SK, Khanna A, Sah RK, Singh I, Raman R. Idiopathic cases of male infertility from a region in India show low incidence of Y-chromosome microdeletion. J Biosci 2003; 28: 605-612.

65. SaoPedro SL, Fraietta R, Spaine D, Porto CS, Srougi M, Cedenho AP, Avellar MC. Prevalence of Y chromosome deletions in a Brazilian population of nonobstructive azoospermic and severely oligozoospermic men. Braz J Med Biol Res 2003; 36: 787-793.

66. Kwok PY. Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet 2001; 2: 235-258.

67. Lee JE, Choi JH, Lee JH, Lee MG. Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. Mutat Res 2005; 573: 195-204.

68. Singh R. RNA-protein interactions that regulate pre-mRNA splicing. Gene Expr 2002; 10: 79-92.

69. Teng YN, Lin YM, Sun HF, Hsu PY, Chung CL, Kuo PL. Association of DAZL haplotypes with spermatogenic failure in infertile men. Fertil Steril 2006; 86: 129-135.

70. Matsushita S, Arai H, Matsui T, Yuzuriha T, Urakami K, Masaki T, Higuchi S. Brain-derived neurotrophic factor gene polymorphisms and Alzheimer's disease. J Neural Transm 2005; 112: 703-711.

71. Iguchi N, Yang S, Lamb DJ, Hecht NB. An SNP in protamine 1: a possible genetic cause of male infertility? J Med Genet 2006; 43: 382-384.

72. Pan SS, Han Y, Farabaugh P, Xia H. Implication of alternative splicing for expression of a variant NAD(P)H:quinone oxidoreductase-1 with a single nucleotide polymorphism at 465C>T. Pharmacogenetics 2002; 12: 479-488.

73. Zuccato E, Buratti E, Stuani C, Baralle FE, Pagani F. An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. J Biol Chem 2004; 279: 16980-16988.

74. Curry BJ, Holt JE, McLaughlin EA, Aitken RJ. Characterization of structure and expression of the Dzip1 gene in the rat and mouse. Genomics 2006; 87: 275-285.

75. Pan HA, Lin YS, Lee KH, Huang JR, Lin YH, Kuo PL. Expression patterns of the DAZ-associated protein DAZAP1 in rat and human ovaries. Fertil Steril 2005; 84 Suppl 2: 1089-1094.

76. Tung JY, Luetjens CM, Wistuba J, Xu EY, Reijo Pera RA, Gromoll J. Evolutionary comparison of the reproductive genes, DAZL and BOULE, in primates with and without DAZ. Dev Genes Evol 2006; 216: 158-168.

77. Urano J, Fox MS, Reijo Pera RA. Interaction of the conserved meiotic regulators, BOULE (BOL) and PUMILIO-2 (PUM2). Mol Reprod Dev 2005; 71: 290-298.

78. Reynolds N, Collier B, Maratou K, Bingham V, Speed RM, Taggart M, Semple CA, Gray NK, Cooke HJ. Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. Hum Mol Genet 2005; 14: 3899-3909.

79. Fox M, Urano J, Reijo Pera RA. Identification and characterization of RNA sequences to which human PUMILIO-2 (PUM2) and deleted in Azoospermia-like (DAZL) bind. Genomics 2005; 85: 92-105.

80. Moore FL, Jaruzelska J, Dorfman DM, Reijo-Pera RA. Identification of a novel gene, DZIP (DAZ-interacting protein), that encodes a protein that interacts with DAZ (deleted in azoospermia) and is expressed in embryonic stem cells and germ cells. Genomics 2004; 83: 834-843.

81. Moore FL, Jaruzelska J, Fox MS, Urano J, Firpo MT, Turek PJ, Dorfman DM, Pera RA. Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. Proc Natl Acad Sci U S A 2003; 100: 538-543.

82. Setzer DR. Measuring equilibrium and kinetic constants using gel retardation assays. Methods Mol Biol 1999; 118: 115-128.

83. Haynes SR. RNA--protein interaction protocols. Totowa, NJ: Humana Press; 1999.

84. Venables JP, Ruggiu M, Cooke HJ. The RNA-binding specificity of the mouse Dazl protein. Nucleic Acids Res 2001; 29: 2479-2483.

85. Dember LM, Kim ND, Liu KQ, Anderson P. Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. Journal of Biological Chemistry 1996; 271: 2783-2788.

86. Baumann C, Otridge J, Gollnick P. Kinetic and thermodynamic analysis of the interaction between TRAP (trp RNA-binding attenuation protein) of *Bacillus subtilis* and trp leader RNA. J Biol Chem 1996; 271: 12269-12274.

87. Tsui S, Dai T, Warren ST, Salido EC, Yen PH. Association of the mouse infertility factor DAZL1 with actively translating polyribosomes. Biol Reprod 2000; 62: 1655-1660.

88. Jiao X, Trifillis P, Kiledjian M. Identification of target messenger RNA substrates for the murine deleted in azoospermia-like RNA-binding protein. Biol Reprod 2002; 66: 475-485.

89. Venables JP, Eperon I. The roles of RNA-binding proteins in spermatogenesis and male infertility. Curr Opin Genet Dev 1999; 9: 346-354.

90. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995; 4: 2411-2423.

91. Dym M, Clermont Y. Role of spermatogonia in the repair of the seminiferous epithelium following x-irradiation of the rat testis. Am J Anat 1970; 128: 265-282.

92. Nakamura M, Okinaga S, Arai K. Metabolism of pachytene primary spermatocytes from rat testes: pyruvate maintenance of adenosine triphosphate level. Biol Reprod 1984; 30: 1187-1197.

93. Gwack Y, Yoo H, Song I, Choe J, Han JH. RNA-Stimulated ATPase and RNA helicase activities and RNA binding domain of hepatitis G virus nonstructural protein 3. J Virol 1999; 73: 2909-2915.

94. Vogel T, Speed RM, Ross A, Cooke HJ. Partial rescue of the Dazl knockout mouse by the human DAZL gene. Mol Hum Reprod 2002; 8: 797-804.

95. Lynch M, Cram DS, Reilly A, O'Bryan MK, Baker HW, de Kretser DM, McLachlan RI. The Y chromosome gr/gr subdeletion is associated with male infertility. Mol Hum Reprod 2005; 11: 507-512.

96. de Vries JW, Hoffer MJ, Repping S, Hoovers JM, Leschot NJ, van der Veen F. Reduced copy number of DAZ genes in subfertile and infertile men. Fertil Steril 2002; 77: 68-75.

97. Maegawa S, Yamashita M, Yasuda K, Inoue K. Zebrafish DAZ-like protein controls translation via the sequence 'GUUC'. Genes Cells 2002; 7: 971-984.

98. Singh K, Raman R. Male infertility: Y-chromosome deletion and testicular aetiology in cases of azoo-/oligospermia. Indian J Exp Biol 2005; 43: 1088-1092.

99. Cheng MH, Maines JZ, Wasserman SA. Biphasic subcellular localization of the DAZL-related protein boule in *Drosophila* spermatogenesis. Developmental Biology (Orlando) 1998; 204: 567-576.

100. Gromoll J, Weinbauer GF, Skaletsky H, Schlatt S, Rocchietti-March M, Page DC, Nieschlag E. The Old World monkey DAZ (Deleted in AZoospermia) gene yields insights into the evolution of the DAZ gene cluster on the human Y chromosome. Hum Mol Genet 1999; 8: 2017-2024.

101. Karashima T, Sugimoto A, Yamamoto M. *Caenorhabditis elegans* homologue of the human azoospermia factor DAZ is required for oogenesis but not for spermatogenesis. Development 2000; 127: 1069-1079.

102. Houston DW, King ML. A critical role for Xdazl, a germ plasm-localized RNA, in the differentiation of primordial germ cells in *Xenopus*. Development 2000; 127: 447-456.

103. Mita K, Yamashita M. Expression of *Xenopus* Daz-like protein during gametogenesis and embryogenesis. Mech Dev 2000; 94: 251-255.

104. Reijo RA, Dorfman DM, Slee R, Renshaw AA, Loughlin KR, Cooke H, Page DC. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. Biol Reprod 2000; 63: 1490-1496.

105. Luetjens CM, Xu EY, Rejo Pera RA, Kamischke A, Nieschlag E, Gromoll J. Association of meiotic arrest with lack of BOULE protein expression in infertile men. J Clin Endocrinol Metab 2004; 89: 1926-1933.

106. Sprague DCC CP, Brown ML, Rice-Ficht AC, and Kuehl TJ. Full-Length DAZ mRNA extracted from mature ejaculated spermatozoa. Biol Reprod 2001, 64 Suppl 1: 219.

107. Vallejo LF, Rinas U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microb Cell Fact 2004; 3: 11.

108. Arraztoa JA, Zhou J, Marcu D, Cheng C, Bonner R, Chen M, Xiang C, Brownstein M, Maisey K, Imarai M, Bondy C. Identification of genes expressed in primate primordial oocytes. Hum Reprod 2005; 20: 476-483.

109. Szczerba A, Jankowska A, Andrusiewicz M, Karczewski M, Turkiewicz W, Warchol JB. Distribution of the DAZ gene transcripts in human testis. Folia Histochem Cytobiol 2004; 42: 119-121.

APPENDICES

Appendix 1. Alignments of DAZL, DAZLmut1, and DAZlmut2

Listed is the published Genbank version of the human DAZL sequence that is aligned with DAZLmut1 and DAZLmut2. There are three lines of nucleotide sequence which are supplemented with additional data above and below their alignments. The amino acid sequence can be found at the beginning of the sequence and shows the TEV protease linker and extends into the beginning of the DAZL gene. The red amino acid notation indicates mutations generated from Teng et al.'s research [15]. The two R's where one is red and one is green indicate the SNP which was found during the course of the current study and is a conservative change.

```
Sequence Analysis
NM_001351 Top Strand
DAZLmut1
DAZLmut2

From-> MBP....
CNGNAAANGCGCAGANTAAT TCG AGC TCG AAC NAC AAC AAC AAT AAC AAT AAC AAC AAC CTC GGG

                                          |-----------TEV-----------|  M   S
                       Glu Phe Ala Leu Gly Glu Asn Leu Tyr Phe Gln Gly Met Ser
ATC GAG GGA AGG ATT TCA GAA TTC GCC CTT GGG GAG AAC CTG TAC TTC CAG GGG ATG TCT
                                          GAG AAC CTG TAC TTC CAG GGG ATG TCT
                                          GAG AAC CTG TAC TTC CAG GGG ATG TCT


 T   A   N   P   E   T   P   N   S   T   I   S   R   E   A   S   T   Q   S   S
Thr Ala Asn Pro Glu Thr Pro Asn Ser Thr Ile Ser Arg Glu Ala Ser Thr Gln Ser Ser
ACT GCA AAT CCT GAA ACT CCA AAC TCA ACC ATC TCC AGA GAG GCC AGC ACC CAG TCC TCA
ACT GCA AAT CCT GAA ACT CCA AAC TCA GCC ATC TCC AGA GAG GCC AGC ACC CAG TCC TCA
                                    Ala
ACT GCA AAT CCT GAA ACT CCA AAC TCA ACC ATC TCC AGA GAG GCC AGC ACC CAG TCC TCA

TCA GCT GCA ACC AGC CAA GGC TAT ATT TTA CCA GAA GGC AAA ATC ATG CCA AAC ACT GTT
TCA GCT GCA ACC AGC CAA GGC TAT ATT TTA CCA GAA GGC AAA ATC ATG CCA AAC ACT GTT
TCA GCT GCA ACC AGC CAA GGC TAT ATT TTA CCA GAA GGC AAA ATC ATG CCA AAC ACT GTT

                                          Thr          R
TTT GTT GGA GGA ATT GAT GTT AGG ATG GAT GAA ACT GAG ATT AGA AGC TTC TTT GCT AGA
TTT GTT GGA GGA ATT GAT GTT AGG ATG GAT GAA ACT GAG ATT AGG AGC TTC TTT GCT AGA
                                                       R
TTT GTT GGA GGA ATT GAT GTT AGG ATG GAT GAA GCT GAG ATT AGG AGC TTC TTT GCT AGA
                                          Ala
```

```
TAT GGT TCA GTG AAA GAA GTG AAG ATA ATC ACT GAT CGA ACT GGT GTG TCC AAA GGC TAT
TAT GGT TCA GTG AAA GAA GTG AAG ATA ATC ACT GAT CGA ACT GGT GTG TCC AAA GGC TAT
TAT GGT TCA GTG AAA GAA GTG AAG ATA ATC ACT GAT CGA ACT GGT GTG TCC AAA GGC TAT

GGA TTT GTT TCA TTT TTT AAT GAC GTG GAT GTG CAG AAG ATA GTA GAA TCA CAG ATA AAT
GGA TTT GTT TCA TTT TTT AAT GAC GTG GAT GTG CAG AAG ATA GTA GAA TCA CAG ATA AAT
GGA TTT GTT TCA TTT TTT AAT GAC GTG GAT GTG CAG AAG ATA GTA GAA TCA CAG ATA AAT

TTC CAT GGT AAA AAG CTG AAG CTG GGC CCT GCA ATC AGG AAA CAA AAT TTA TGT GCT TAT
TTC CAT GGT AAA AAG CTG AAG CTG GGC CCT GCA ATC AGG AAA CAA AAT TTA TGT GCT TAT
TTC CAT GGT AAA AAG CTG AAG CTG GGC CCT GCA ATC AGG AAA CAA AAT TTA TGT GCT TAT

CAT GTG CAG CCA CGT CCT TTG GTT TTT AAT CAT CCT CCT CCA CCA CAG TTT CAG AAT GTC
CAT GTG CAG CCA CGT CCT TTG GTT TTT AAT CAT CCT CCT CCA CCA CAG TTT CAG AAT GTC
CAT GTG CAG CCA CGT CCT TTG GTT TTT AAT CAT CCT CCT CCA CCA CAG TTT CAG AAT GTC

TGG ACT AAT CCA AAC ACT GAA ACT TAT ATG CAG CCC ACA ACC ACG ATG AAT CCT ATA ACT
TGG ACT AAT CCA AAC ACT GAA ACT TAT ATG CAG CCC ACA ACC ACG ATG AAT CCT ATA ACT
TGG ACT AAT CCA AAC ACT GAA ACT TAT ATG CAG CCC ACA ACC ACG ATG AAT CCT ATA ACT



CAG TAT GTT CAG GCA TAT CCT ACT TAC CCA AAT TCA CCA GTT CAG GTC ATC ACT GGA TAT
CAG TAT GTT CAG GCA TAT CCT ACT TAC CCA AAT TCA CCA GTT CAG GTC ATC ACT GGA TAT
CAG TAT GTT CAG GCA TAT CCT ACT TAC CCA AAT TCA CCA GTT CAG GTC ATC ACT GGA TAT


CAG TTG CCT GTA TAT AAT TAT CAG ATG CCA CCA CAG TGG CCT GTT GGG GAG CAA AGG AGC
CAG TTG CCT GTA TAT AAT TAT CAG ATG CCA CCA CAG TGG CCT GTT GGG GAG CAA AGG AGC
CAG TTG CCT GTA TAT AAT TAT CAG ATG CCA CCA CAG TGG CCT GTT GGG GAG CAA AGG AGC

TAT GTT GTA CCT CCG GCT TAT TCA GCT GTT AAC TAC CAC TGT AAT GAA GTT GAT CCA GGA
TAT GTT GTA CCT CCG GCT TAT TCA GCT GTT AAC TAC CAC TGT AAT GAA GTT GAT CCA GGA
TAT GTT GTA CCT CCG GCT TAT TCA GCT GTT AAC TAC CAC TGT AAT GAA GTT GAT CCA GGA

GCT GAA GTT GTG CCA AAT GAA TGT TCA GTT CAT GAA GCT ACT CCA CCC TCT GGA AAT GGC
GCT GAA GTT GTG CCA AAT GAA TGT TCA GTT CAT GAA GCT ACT CCA CCC TCT GGA AAT GGC
GCT GAA GTT GTG CCA AAT GAA TGT TCA GTT CAT GAA GCT ACT CCA CCC TCT GGA AAT GGC

CCA CAA AAG AAA TCT GTG GAC CGA AGC ATA CAA ACG GTG GTA TCT TGT CTG TTT AAT CCA
CCA CAA AAG AAA TCT GTG GAC CGA AGC ATA CAA ACG GTG GTA TCT TGT CTG TTT AAT CCA
CCA CAA AAG AAA TCT GTG GAC CGA AGC ATA CAA ACG GTG GTA TCT TGT CTG TTT AAT CCA

GAG AAC AGA CTG AGA AAC TCT GTT GTT ACT CAA GAT GAC TAC TTC AAG GAT AAA AGA GTG
GAG AAC AGA CTG AGA AAC TCT GTT GTT ACT CAA GAT GAC TAC TTC AAG GAT AAA AGA GTG
GAG AAC AGA CTG AGA AAC TCT GTT GTT ACT CAA GAT GAC TAC TTC AAG GAT AAA AGA GTG

                                                  Stop|
CAT CAC TTT AGA AGA AGT CGG GCA ATG CTT AAA TCT GTT TGA
CAT CAC TTT AGA AGA AGT CGG GCA ATG CTT AAA TCT GTT TGA
CAT CAC TTT AGA AGA AGT CGG GCA ATG CTT AAA TCT GTT TGA
```

Appendix 2.Comparison of NM_020363 (DAZ2) with Cloned Version of DAZ

Two different DAZ sequences are aligned with the DAZ sequence cloned for use in the current study. The main feature of note is the makeup of the repeat region of the DAZ gene cloned for the current study. Following the sequences are the sequences of the repeat regions and their comparisons with each other.

```
NM_020363, DAZ2, Transcript Variant 1 Top Strand
Cloned DAZ Sequence, 2nd Strand
NM_020364, DAZ3, 3rd Strand
Repeat Region Identifications from Saxena, Nature Genetics, Vol 14, November 1996, pg 292

E1 |Exon 2 --->
ATG|TCTGCTGCAAATCCTGAGACTCCAAACTCAACCATCTCCAGAGAGGCCAGCACCCAGTCTTCATCAGCTGCAGCTAGCCAAGGCT
ATG|TCTGCTGCAAATCCTGAGACTCCAAACTCAACCATCTCCAGAGAGGCCAGCACCCAGTCTTCATCAGCTGCAGCTAGCCAAGGCT
ATG|TCTGCTGCAAATCCTGAGACTCCAAACTCAACCATCTCCAGAGAGGCCAGCACCCAGTCTTCATCAGCTGCAGCTAGCCAAGGCT


                                                                |Exon 3 --->
GGGTGTTACCAGAAGGCAAAATCGTGCCAAACACTGTTTTTGTTGGTGGAATTGATGCTAGG|ATGGATGAAACTGAGATTGGAAGCTG
GGGTGTTACCAGAAGGCAAAATCGTGCCAAACACTGTTTTTGTTGGTGGAATTGATGCTAGG|ATGGATGAAACTGAGATTGGAAGCTG
GGGTGTTACCAGAAGGCAAAATCGTGCCAAACACTGTTTTTGTTGGTGGAATTGATGCTAGG|ATGGATGAAACTGAGATTGGAAGCTG


                                                                  |Exon 4 --->
CTTTGGTAGATACGGTTCAGTGAAAGAAGTGAAGATAATCACGAATCGAACTGGTGTGTCCAAAGG|CTATGGATTTGTTTCGTTTGTT
CTTTGGTAGATACGGTTCAGTGAAAGAAGTGAAGATAATCACGAATCGAACTGGTGTGTCCAAAGG|CTATGGATTTGTTTCGTTTGTT
CTTTGGTAGATACGGTTCAGTGAAAGAAGTGAAGATAATCACGAATCGAACTGGTGTGTCCAAAGG|CTATGGATTTGTTTCGTTTGTT


                                |Exon 5 --->
AATGACGTGGATGTCCAGAAGATAGTAGGA|TCACAGATACATTTCCATGGTAAAAAGCTGAAGCTGGGCCCTGCAATCAGGAAACAAA
AATGACGTGGATGTCCAGAAGATAGTAGGA|TCACAGATACATTTCCATGGTAAAAAGCTGAAGCTGGGCCCTGCAATCAGGAAACAAA
AATGACGTGGATGTCCAGAAGATAGTAGGA|TCACAGATACATTTCCATGGTAAAAAGCTGAAGCTGGGCCCTGCAATCAGGAAACAAA




      |Exon 6 --->
AGTTAT|GTGCTCGTCATGTGCAGCCACGTCCTTTGGTAGTTAATCCTCCTCCTCCACCACAGTTTCAGAACGTCTGGCGGAATCCAAA
AGTTAT|GTGCTCGTCATGTGCAGCCACGTCCTTTGGTAGTTAATCCTCCTCCTCCACCACAGTTTCAGAACGTCTGGCGGAATCCAAA
AGTTAT|GTGCTCGTCATGTGCAGCCACGTCCTTTGGTAGTTAATCCTCCTCCTCCACCACAGTTTCAGAACGTCTGGCGGAATCCAAA


                                                              |Exon 7a --->
CACTGAAACCTACCTGCAGCCCCAAATCACGCCGAATCCTGTAACTCAGCACGTTCAG|GCTTATTCTGCTTATCCACATTCACCAGGT
CACTGAAACCTACCTGCAGCCCCAAATCACGCCGAATCCTGTAACTCAGCACGTTCAG|GCTTACTCTGCTTATCCACATTCACCAGGT
CACTGAAACCTACCTGCAGCCCCAAATCACGCCGAATCCTGTAACTCAGCACGTTCAG|


                                          |Exon 7b --->
CAGGTCATCACTGGATGTCAGTTGCTTGTATATAATTATCAG|GAATATCCTACTTATCCCGATTCAGCATTTCAGGTCACCACTGGAT
CAGGTCATCACTGGATGTCAGTTGCTTGTATATAATTATCAG|GAATATCCTACTTATCCCGATTCAGCATTTCAGGTCACCACTGGAT
```

```
                          |Exon 7c --->
ATCAGTTGCCTGTATATAATTATCAG|CCATTTCCTGCTTATCCAAGATCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA
ATCAGTTGCCTGTATATAATTATCAG|CCATTTCCTGCTTATCCAAAATCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA


         |Exon 7d --->                                                                |Exon
TAATTATCAG|GCATTTCCTGCTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATACAATTATCAG|CCATT
TAATTATCAG|CCATTTCCTGCTTATCCAAGATCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATATAATTATCAG|GCATT


7e ---->                                                              |Exon 7f --->
TCCTGCTTATCCAAGTTCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATATAATTATCAG|GCATTTCCTGCTTATCCAAAT
TCCTGCTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATACAATTATCAG|CCATTTCCTGCTTATCCAAGT

                                                    |Exon 7g --->
TCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATACAATTATCAG|GCATTTCCTGCTTATCCAAATTCACCAGTTCAGGTCA
TCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATATAATTATCAG|

                                  |Exon 7h --->
CCACTGGATATCAGTTGCCTGTATACAATTATCAG|GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTT

                   |Exon 7i --->
GCCTGTATATAATTATCAG|GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTAT

   |Exon 7j --->                                                                  |Exon 7k --->
CAG|GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAG|GCATTTCCTGCT

                                                          |Exon 7l --->
TATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAG|GCATTTCCTGCTTATCCAAGTTCACCAT

                                           |Exon 7m --->
TTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAG|GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGG

                             |Exon 7n --->
ATATCAGTTGCCTGTATATAATTATCAG|GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTA

            |Exon 7o --->                                                              |Exo
TATAATTATCAG|GCATTTCCTGCTTATCCAAATTCAGCAGTTCAGGTCACCACTGGATATCAGTTCCATGTATACAATTACCAG|ATG
                                                                                       |ATG
                                                                                       |ATG


n 8 --->                         |Exon 10 ---->
CCACCGCAGTGCCCTGTTGGGGAGCAAAGGAG|AAATCTGTGGACCGAAGCATACAAATGGTGGTATCTTGTCTGTTTAATCCAGAGAA
CCACCGCAGTGCCCTGTTGGGGAGCAAAGGAG|AAATCTGTGGACCGAAGCATACAAATGGTGGTATCTTGTCTGTTTAATCCAGAGAA
CCACCGCAGTGCCCTGTTGGGGAGCAAAGGAG|AAATCTGTGGACCGAAGCATACAAATGGTGGTATCTTGTCTGTTTAATCCAGAGAA


GAGACTGA
GAGACTGA
GAGACTGA



DAZX Repeat Region- Total Number of Repeats=10

Exon 7a
GCT TAC TCT G CTTATCCACATTCACCAGGTCAGGTCATCACTGGATGTCAGTTGCTTGTATA TAA TTA TCA G

Exon 7b
GAA TAT CCT A CTTATCCCGATTCAGCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7c
CCA TTT CCT G CTTATCCAAAATCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA TAA TTA TCA G
```

Exon 7c
CCA TTT CCT G CTTATCCAAGATCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATA CAA TTA TCA G

Exon 7c
CCA TTT CCT G CTTATCCAAGTTCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCAGTTCAGGTCACCACTGGATATCAGTTGCCTGTATA CAA TTA TCA G

Exon 7f
GCA TTT CCT G CTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7g
GCA TTT CCT G CTTATCCAAATTCAGCAGTTCAGGTCACCACTGGATATCAGTTCCATGTATA CAA TTA CCA G

NM_020364 (DAZ3) Repeat Region- Total Number of Repeats=12

Exon 7a
GCT TAC TCT G CTTATCCACATTCACCAGGTCAGGTCATCACTGGATGTCAGTTGCTTGTATA TAA TTA TCA G

Exon 7b
GAA TAT CCT A CTTATCCCGATTCAGCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7c
CCA TTT CCT G CTTATCCAAGATCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATA CAA TTA TCA G

Exon 7C
CCA TTT CCT G CTTATCCAAGTTCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATA TAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCAGTTCAGGTCACCACTGGATATCAGTTGCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCAGTTCAGGTCACCACTGGATATCAGTTGCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCATTTCAAGTCGCCACTGGATATCAGTTCCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCACCAGTTCAGGTCACCACTGGATATCAGTTGCCTGTATA CAA TTA TCA G

Exon 7d
GCA TTT CCT G CTTATCCAAATTCAGCAGTTCAGGTCACCACTGGATATCAGTTCCATGTATA CAA TTA CCA G

Appendix 3. Concatenation Scheme- Nucleotide Level

The nucleotide level of the concatenation scheme showing the final product produced is shown below. The first round of PCR creates a shortened sequence that is extended upon further PCR cycles. 1 indicates the complete product from the PCR reactions. 2 shows the result of the first round of PCR and 3 shows the completion of the second round of PCR.

1- This is the complete product of the PCR. Bold indicates ordered primers. Underlined indicates the ordered "template oligo." Full product size is 121bp.

T7 Promoter-containing primer

5′ **CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT** GTTTTTTTGTTTTTTTTTTTGTTTTTTTTTTTGTTTTTTT GTT GAT ATA AAT AGT GCC CAT GGA TCC GCG GGT GTC GGG 3′

3′ GCG CCT AGG ATT ATG CTG AGT GAT ATC CCC GGT GGT TGC TGT AA CAAAAAAACAAAAAAAAAACAAAAAAAAAACAAAAAAA **CAA CTA TAT TTA TCA CGG GTA CCT AGG CGC CCA CAG CCC** 5′

CDC25C Oligo Rev. Univ. Primer

2- First round of PCR shows the forward primer extending its 3′ end to the end of the template oligo and the extension of the 3′ end of the template oligo.

5′ **CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT** 3′

3′ CCC GGT GGT TGC TGT AA CAAAAAAACAAAAAAAAAACAAAAAAAAAACAAAAAAA **CAA CTA TAT TTA TCA CGG GT** 5′

Product of first and subsequent rounds

5′ **CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT** GTTTTTTTGTTTTTTTTTTTGTTTTTTTTTTTGTTTTTTT GTT GAT ATA AAT AGT GCC CA 3′

3′ GCG CCT AGG ATT ATG CTG AGT GAT ATC CCC GGT GGT TGC TGT AA CAAAAAAACAAAAAAAAAACAAAAAAAAAACAAAAAAA CAA CTA TAT TTA TCA CGG GT 5′

3- Second round of PCR shows the completion of the whole PCR product

5′ **CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT** GTTTTTTTGTTTTTTTTTTTGTTTTTTTTTTTGTTTTTTT GTT GAT ATA AAT AGT GCC CA 3′

3′ **CAA CTA TAT TTA TCA CGG GTA CCT AGG CGC CCA CAG CCC** 5′

Product of second and subsequent rounds

5′ **CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT** GTTTTTTTGTTTTTTTTTTTGTTTTTTTTTTTGTTTTTTT GTT GAT ATA AAT AGT GCC CAT GGA TCC GCG GGT GTC GGG 3′

3′ GCG CCT AGG ATT ATG CTG AGT GAT ATC CCC GGT GGT TGC TGT AA CAAAAAAACAAAAAAAAAACAAAAAAAAAACAAAAAAA **CAA CTA TAT TTA TCA CGG GTA CCT AGG CGC CCA CAG CCC** 5′

Primers and oligos in reaction

5′ **CGC GGA TCC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT** 3′

3′ CCC GGT GGT TGC TGT AA CAAAAAAACAAAAAAAAAACAAAAAAAAAACAAAAAAA **CAA CTA TAT TTA TCA CGG GT** 5′

3′ **CAA CTA TAT TTA TCA CGG GTA CCT AGG CGC CCA CAG CCC** 5′

4- A gel of PCR products should show a primary band at 121 bp without detectable bands at 103 and 94 bp representing incomplete construction.

Appendix 4. Three-Fold or Greater Targets of DAZL Binding

The following table lists in order of fold increase over MBP and with p-values of less than $p < 0.05$ the sequences selected by DAZL in the microarray study. Only the sequences of 3-fold and greater are shown. The column lists Systematic ID which is used by the GeneSpring software, RefSeq used by Genbank, the name of the gene or product, the process in which the gene is involved, the normalized number, the ratio, and the p-value.

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_014814	NM_014814	proteasome regulatory particle subunit p44S10	ATP-dependent proteolysis	0.0703	14.22	0.0024
NM_000019	NM_000019	acetyl-Coenzyme A acetyltransferase 1 precursor		0.0731	13.67	0.0012
AK057433	NM_144674	hypothetical protein FLJ32871	microtubule cytoskeleton organization and biogenesis	0.0784	12.76	0.0090
NM_005977	NM_005977; NM_183043; NM_183044; NM_183045	ring finger protein 6 isoform 1; ring finger protein 6 isoform 2		0.0811	12.33	0.0011
NM_006197	NM_006197	pericentriolar material 1		0.0812	12.31	0.0017
NM_004872	NM_004872	thymic dendritic cell-derived factor 1		0.0835	11.97	0.0012
NM_014932	NM_014932	neuroligin 1	synaptogenesis; protein targeting; cell adhesion; synaptic vesicle targeting; calcium-dependent cell-cell adhesion; ion channel clustering; regulation of neuron differentiation	0.0853	11.73	0.0013
NM_007051	NM_007051; NM_131917	FAS-associated factor 1 isoform a; FAS-associated factor 1 isoform b	apoptosis	0.0870	11.50	0.0010
NM_007204	NM_007204	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20	assembly of spliceosomal tri-snRNP; mRNA processing	0.0889	11.25	0.0023
AK056446	NM_005348	heat shock 90kDa protein 1, alpha		0.0971	10.30	0.0038
NM_030933	NM_030933	chromosome 1 open reading frame 14		0.0981	10.19	0.0039
NM_014633	NM_014633	SH2 domain binding protein 1	regulation of transcription, DNA-dependent	0.0982	10.18	0.0041
NM_003295	NM_003295	tumor protein, translationally-controlled 1		0.1052	9.51	0.0057
NM_001006	NM_001006; NM_182777	ribosomal protein S3a	protein biosynthesis	0.1061	9.43	0.0074
BC017296	NM_144665	sestrin 3	cell cycle arrest	0.1075	9.30	0.0016
NM_006554	NM_006554	metaxin 2	mitochondrial transport; intracellular protein transport	0.1089	9.18	0.0046
NM_003299	NM_003299	tumor rejection antigen (gp96) 1	response to stress; protein folding	0.1096	9.12	0.0031
NM_006461	NM_006461	mitotic spindle coiled-coil related protein		0.1097	9.12	0.0063
NM_002592	NM_002592; NM_182649	proliferating cell nuclear antigen	DNA repair; DNA replication; cell proliferation; cell cycle	0.1109	9.01	0.0025

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
			control			
NM_053023	NM_053023; NM_170768	zinc finger protein 91 isoform 1; zinc finger protein 91 isoform 2		0.1109	9.01	0.0306
NM_003134	NM_003134	signal recognition particle 14kDa (homologous Alu RNA binding protein)	protein targeting; cotranslational membrane targeting	0.1125	8.89	0.0027
NM_004559	NM_004559	nuclease sensitive element binding protein 1	response to pest/pathogen/parasite; transcription from Pol II promoter	0.1126	8.88	0.0054
NM_030941	NM_030941	exonuclease NEF-sp		0.1135	8.81	0.0062
AB051469	NM_019061	phosphatidylinositol-3 phosphate 3-phosphatase adaptor subunit		0.1151	8.69	0.0020
NM_006584	NM_006584	chaperonin containing TCP1, subunit 6B (zeta 2)	protein complex assembly, multichaperone pathway; spermatogenesis; protein folding	0.1169	8.55	0.0055
NM_016027	NM_016027	lactamase, beta 2		0.1173	8.53	0.0012
NM_003447	NM_003447	zinc finger protein 165	regulation of transcription, DNA-dependent	0.1174	8.52	0.0016
Z68274				0.1186	8.43	0.0036
NM_003296	NM_003296	testis specific protein 1		0.1195	8.37	0.0062
NM_006430	NM_006430	chaperonin containing TCP1, subunit 4 (delta)	protein folding; regulation of cell cycle	0.1196	8.36	0.0012
AL050335				0.1203	8.31	0.0030
NM_032631	NM_032631	hepatoma-derived growth factor-related protein 2	regulation of transcription, DNA-dependent	0.1210	8.27	0.0024
NM_006835	NM_006835	cyclin I	spermatogenesis	0.1230	8.13	0.0054
AJ224082				0.1231	8.13	0.0018
NM_019644	NM_019644	testis-specific ankyrin motif containing protein	male gonad development	0.1232	8.12	0.0009
AK057303	NM_139243	testis nuclear RNA-binding protein	RNA processing; proteolysis and peptidolysis	0.1238	8.07	0.0018
NM_000971	NM_000971	ribosomal protein L7	regulation of transcription, DNA-dependent; protein biosynthesis	0.1242	8.05	0.0063
NM_005192	NM_005192	cyclin-dependent kinase inhibitor 3	G1/S transition of mitotic cell cycle; cell cycle arrest; negative regulation of cell proliferation; regulation of CDK activity	0.1244	8.04	0.0057
NM_002948	NM_002948	ribosomal protein L15	protein biosynthesis	0.1248	8.01	0.0017
U93181	NM_002972	SET binding factor 1 isoform a	protein amino acid dephosphorylation	0.1265	7.90	0.0019
NM_001015	NM_001015	ribosomal protein S11	protein biosynthesis	0.1266	7.90	0.0069
NM_003756	NM_003756	eukaryotic translation initiation factor 3, subunit 3 gamma, 40kDa	regulation of translational initiation; protein biosynthesis	0.1279	7.82	0.0020
NM_020366	NM_020366	retinitis pigmentosa GTPase regulator interacting protein 1		0.1284	7.79	0.0039
NM_032135	NM_032135	hypothetical protein DKFZp434F1017		0.1287	7.77	0.0033
AL049690				0.1289	7.76	0.0026
NM_017903	NM_017903	hypothetical protein FLJ20618		0.1290	7.75	0.0017
NM_022443	NM_022443	myeloid leukemia factor 1		0.1291	7.75	0.0035
NM_018473	NM_018473	thioesterase superfamily member 2		0.1298	7.70	0.0042
BC014595	NM_145046	calreticulin 3		0.1311	7.63	0.0023
NM_001005	NM_001005	ribosomal protein S3	protein biosynthesis	0.1314	7.61	0.0039
NM_017917	NM_017917	chromosome 14 open reading frame 10		0.1323	7.56	0.0184

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_015571	NM_015571	SUMO-1-specific protease	proteolysis and peptidolysis	0.1334	7.49	0.0070
AK055354	NM_005131	nuclear matrix protein p84	transport; mRNA-nucleus export; signal transduction; nuclear mRNA splicing, via spliceosome	0.1339	7.47	0.0040
NM_032580	NM_032580	hairy and enhancer of split 7	mesoderm development; regulation of transcription, DNA-dependent	0.1340	7.46	0.0032
NM_018686	NM_018686	cytidine 5'-monophosphate N-acetylneuraminic acid synthetase		0.1341	7.46	0.0036
NM_000183	NM_000183	hydroxyacyl dehydrogenase, subunit B	fatty acid beta-oxidation; fatty acid metabolism	0.1343	7.44	0.0021
NM_002631	NM_002631	phosphogluconate dehydrogenase	pentose-phosphate shunt, oxidative branch	0.1344	7.44	0.0099
NM_006837	NM_006837	COP9 signalosome subunit 5	protein biosynthesis; transcription from Pol II promoter	0.1347	7.42	0.0061
NM_001418	NM_001418	eukaryotic translation initiation factor 4 gamma, 2	cell cycle arrest; cell death; regulation of translational initiation; regulation of protein biosynthesis	0.1348	7.42	0.0050
BC010112	NM_002156	heat shock 60kDa protein 1 (chaperonin)		0.1351	7.40	0.0038
NM_004505	NM_004505	ubiquitin specific protease 6	deubiquitination; oncogenesis; protein modification	0.1355	7.38	0.0081
NM_002106	NM_002106	H2A histone family, member Z	nucleosome assembly; chromosome organization and biogenesis (sensu Eukarya)	0.1359	7.36	0.0083
NM_032315	NM_032315	mitochondrial carrier protein MGC4399	transport	0.1359	7.36	0.0018
NM_006754	NM_006754; NM_182715	synaptophysin-like protein isoform a; synaptophysin-like protein isoform b	transport; nonselective vesicle transport; synaptic transmission	0.1362	7.34	0.0039
NM_001351	NM_001351	deleted in azoospermia-like	spermatogenesis	0.1366	7.32	0.0064
NM_003992	NM_001292; NM_003992	CDC-like kinase 3 isoform hclk3/152	protein amino acid phosphorylation	0.1377	7.26	0.0028
NM_001749	NM_001749	calpain, small subunit 1	positive regulation of cell proliferation	0.1378	7.26	0.0043
AL096829				0.1378	7.25	0.0037
NM_016930	NM_016930	syntaxin 18	ER to Golgi transport; nonselective vesicle docking; intracellular protein transport	0.1378	7.25	0.0085
BC007261		Unknown (protein for MGC:15545)		0.1382	7.23	0.0042
NM_002787	NM_002787	proteasome alpha 2 subunit	ubiquitin-dependent protein catabolism	0.1386	7.22	0.0029
NM_025146	NM_025146	hypothetical protein FLJ13194		0.1386	7.22	0.0051
NM_017425	NM_017425	sperm autoantigenic protein 17	spermatogenesis; fertilization (sensu Animalia); signal transduction	0.1387	7.21	0.0079
NM_004902	NM_004902; NM_184234; NM_184237; NM_184241; NM_184244	RNA-binding region containing protein 2 isoform b; RNA-binding region containing protein 2 isoform a; RNA-binding region containing protein 2 isoform c; RNA-binding region containing protein 2 isoform d	regulation of transcription, DNA-dependent; nuclear mRNA splicing, via spliceosome	0.1404	7.12	0.0032
NM_002687	NM_002687	pinin, desmosome associated protein	cell adhesion; cell shape and cell size control	0.1408	7.10	0.0352
NM_014056	NM_014056	DKFZP564K247 protein		0.1411	7.09	0.0010
NM_006029	NM_006029	paraneoplastic antigen MA1	spermatogenesis; central nervous system development	0.1412	7.08	0.0052

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_003350	NM_003350	ubiquitin-conjugating enzyme E2 variant 2	protein polyubiquitination; regulation of DNA repair; ubiquitin cycle; cell proliferation; regulation of cell cycle	0.1413	7.08	0.0064
NM_002790	NM_002790	proteasome alpha 5 subunit	ubiquitin-dependent protein catabolism	0.1414	7.07	0.0008
NM_032858	NM_032858	hypothetical protein FLJ14904		0.1422	7.03	0.0037
NM_005493	NM_005493	RAN binding protein 9	microtubule nucleation; cell growth and/or maintenance; protein complex assembly	0.1426	7.01	0.0046
BC008767	NM_004035; NM_007292	acyl-Coenzyme A oxidase isoform a; acyl-Coenzyme A oxidase isoform b	fatty acid beta-oxidation; electron transport; prostaglandin metabolism; energy pathways	0.1426	7.01	0.0008
NM_002520	NM_002520; NM_199185	nucleophosmin 1		0.1429	7.00	0.0022
NM_005154	NM_005154	ubiquitin specific protease 8	ubiquitin-dependent protein catabolism; cell proliferation	0.1435	6.97	0.0013
NM_014362	NM_014362; NM_198047	3-hydroxyisobutyryl-Coenzyme A hydrolase isoform 1; 3-hydroxyisobutyryl-Coenzyme A hydrolase isoform 2	metabolism	0.1442	6.94	0.0102
NM_004986	NM_004986; NM_182926	kinectin 1	non-selective vesicle transport	0.1443	6.93	0.0054
NM_004456	NM_004456; NM_152998	enhancer of zeste 2 isoform a; enhancer of zeste 2 isoform b	chromatin architecture; transcription regulation	0.1444	6.93	0.0029
NM_003368	NM_003368	ubiquitin specific protease 1	ubiquitin-dependent protein catabolism	0.1451	6.89	0.0007
NM_030939	NM_030939	chromosome 6 open reading frame 62		0.1451	6.89	0.0045
NM_021190	NM_021190	polypyrimidine tract binding protein 2		0.1455	6.87	0.0009
AK057501	NM_144673	chemokine-like factor super family 2		0.1469	6.81	0.0166
BC010281	NM_015161	ADP-ribosylation factor-like 6 interacting protein		0.1472	6.79	0.0001
NM_006016	NM_006016	CD164 antigen, sialomucin	negative regulation of cell proliferation; cell adhesion; development; immune response; signal transduction	0.1478	6.76	0.0204
NM_032905	NM_032905	splicing factor (45kD)	mRNA processing; mRNA splicing; nuclear mRNA splicing, via spliceosome	0.1482	6.75	0.0063
BC014598	NM_145269	similar to CG6405 gene product		0.1483	6.74	0.0027
NM_016091	NM_016091	eukaryotic translation initiation factor 3, subunit 6 interacting protein	protein biosynthesis	0.1488	6.72	0.0006
NM_004786	NM_004786	thioredoxin-like, 32kDa	protein-disulfide reduction; apoptosis; signal transduction	0.1490	6.71	0.0016
AK023096				0.1491	6.71	0.0473
NM_005998	NM_005998	chaperonin containing TCP1, subunit 3 (gamma)	protein folding	0.1505	6.65	0.0020
AK058062	NM_152548	hypothetical protein FLJ25333		0.1508	6.63	0.0023
NM_031243	NM_002137; NM_031243	heterogeneous nuclear ribonucleoprotein A2/B1 isoform A2; heterogeneous nuclear ribonucleoprotein A2/B1 isoform B1	RNA processing	0.1513	6.61	0.0065
NM_006406	NM_006406	thioredoxin peroxidase	I-kappaB phosphorylation	0.1518	6.59	0.0047
NM_016220	NM_016220	zinc finger protein (ZFD25)	regulation of transcription, DNA-dependent; biological_process unknown	0.1522	6.57	0.0014

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_001831	NM_001831	clusterin	complement activation, classical pathway; cell death; lipid metabolism; fertilization (sensu Animalia); apoptosis	0.1524	6.56	0.0086
NM_030752	NM_030752	t-complex 1	protein folding	0.1526	6.55	0.0051
NM_012189	NM_012189; NM_138643; NM_138644; NM_153768; NM_153769; NM_153770	calcium-binding tyrosine phosphorylation-regulated protein isoform a; calcium-binding tyrosine phosphorylation-regulated protein isoform d; calcium-binding tyrosine phosphorylation-regulated protein isoform c; calcium-binding tyrosine phosphorylation-regulated protein isoform b; calcium-binding tyrosine phosphorylation-regulated protein isoform e		0.1526	6.55	0.0087
NM_001726	NM_001726	testis-specific bromodomain protein		0.1528	6.54	0.0016
NM_033546	NM_033546	myosin regulatory light chain MRCL2		0.1533	6.52	0.0048
NM_002799	NM_002799	proteasome beta 7 subunit proprotein	ubiquitin-dependent protein catabolism	0.1540	6.49	0.0100
NM_022807	NM_003097; NM_022805; NM_022806; NM_022807; NM_022808	small nuclear ribonucleoprotein polypeptide N	mRNA splicing	0.1541	6.49	0.0023
NM_004238	NM_004238	thyroid hormone receptor interactor 12	protein ubiquitination	0.1541	6.49	0.0082
NM_002497	NM_002497	NIMA (never in mitosis gene a)-related kinase 2	regulation of mitosis; meiosis; protein amino acid phosphorylation; cytokinesis	0.1552	6.44	0.0065
AK055714	NM_002789	proteasome alpha 4 subunit	ubiquitin-dependent protein catabolism	0.1553	6.44	0.0018
NM_002539	NM_002539	ornithine decarboxylase 1	polyamine biosynthesis	0.1553	6.44	0.0101
NM_016640	NM_016640	mitochondrial ribosomal protein S30	apoptosis	0.1556	6.43	0.0010
NM_001359	NM_001359	2,4-dienoyl CoA reductase 1 precursor	metabolism	0.1559	6.42	0.0025
NM_018093	NM_018093	hypothetical protein FLJ10439		0.1561	6.41	0.0029
NM_005336	NM_005336	high density lipoprotein binding protein	lipid transport; cholesterol metabolism	0.1565	6.39	0.0109
NM_006716	NM_006716	activator of S phase kinase	DNA replication; cell cycle control; mitotic G1/S transition	0.1566	6.39	0.0023
AB033079	NM_020755	tumor differentially expressed 2		0.1569	6.37	0.0058
NM_001320	NM_001320	casein kinase 2, beta polypeptide	signal transduction	0.1571	6.37	0.0037
NM_003143	NM_003143	single-stranded DNA binding protein 1	DNA replication	0.1572	6.36	0.0092
NM_017743	NM_017743; NM_130434; NM_197960; NM_197961	dipeptidyl peptidase 8 isoform 2; dipeptidyl peptidase 8 isoform 1; dipeptidyl peptidase 8 isoform 3; dipeptidyl peptidase 8 isoform 4		0.1572	6.36	0.0029
NM_032559	NM_032559	kinesin protein		0.1574	6.35	0.0039

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
BC017117	NM_001881; NM_181571; NM_182717; NM_182718; NM_182719; NM_182720; NM_182721; NM_182722; NM_182723; NM_182724; NM_182725; NM_182769; NM_182770; NM_182771; NM_182772; NM_182850; NM_182853; NM_183011; NM_183012; NM_183013; NM_183060	cAMP responsive element modulator isoform b; cAMP responsive element modulator isoform a; cAMP responsive element modulator isoform d; cAMP responsive element modulator isoform e; cAMP responsive element modulator isoform f; cAMP responsive element modulator isoform g; cAMP responsive element modulator isoform h; cAMP responsive element modulator isoform i; cAMP responsive element modulator isoform j; cAMP responsive element modulator isoform k; cAMP responsive element modulator isoform l; cAMP responsive element modulator isoform m; cAMP responsive element modulator isoform n; cAMP responsive element modulator isoform o; cAMP responsive element modulator isoform p; cAMP responsive element modulator isoform q; cAMP responsive element modulator isoform r; cAMP responsive element modulator isoform u; cAMP responsive element modulator isoform t; cAMP responsive element modulator isoform s; cAMP responsive element modulator isoform v	signal transduction	0.1576	6.35	0.0031
NM_002085	NM_002085	glutathione peroxidase 4	phospholipid metabolism; response to oxidative stress; development	0.1579	6.33	0.0104
D38553	NM_015341	barren	mitotic chromosome condensation; cell cycle; mitosis	0.1580	6.33	0.0108
NM_006667	NM_006667	progesterone receptor membrane component 1		0.1581	6.33	0.0001
NM_004881	NM_004881; NM_147184	tumor protein p53 inducible protein 3		0.1582	6.32	0.0020
AK055287				0.1583	6.32	0.0042
NM_003973	NM_003973	ribosomal protein L14	protein biosynthesis	0.1586	6.31	0.0053
NM_003093	NM_003093	small nuclear ribonucleoprotein polypeptide C	RNA splicing	0.1590	6.29	0.0031
NM_006007	NM_006007	zinc finger protein 216	biological_process unknown	0.1595	6.27	0.0043
AL136939	NM_021814	homolog of yeast long chain polyunsaturated fatty acid elongatio		0.1596	6.27	0.0025
NM_018180	NM_018180	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 32		0.1597	6.26	0.0001
NM_006265	NM_006265	RAD21 homolog	meiotic recombination; chromosome segregation; cell cycle; mitosis; double-strand break repair; apoptosis	0.1603	6.24	0.0003
NM_021794	NM_021794	a disintegrin and metalloproteinase domain 30 preproprotein	proteolysis and peptidolysis	0.1610	6.21	0.0016
NM_006768	NM_006768	BRCA1 associated protein	nucleocytoplasmic transport	0.1611	6.21	0.0010
NM_002950	NM_002950	ribophorin I	protein modification	0.1613	6.20	0.0047
NM_002893	NM_002893	retinoblastoma binding protein 7	development; cell proliferation	0.1616	6.19	0.0006
AF132027				0.1618	6.18	0.0035
NM_030979	NM_030979	poly(A) binding protein, cytoplasmic 3	mRNA metabolism	0.1619	6.18	0.0065
NM_033122	NM_033122	testis development protein NYD-SP26		0.1623	6.16	0.0038

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_001564	NM_001564	inhibitor of growth family, member 1-like	regulation of transcription, DNA-dependent; signal transduction	0.1624	6.16	0.0040
NM_003896	NM_003896	sialyltransferase 9	protein amino acid glycosylation	0.1627	6.15	0.0019
NM_001920	NM_001920; NM_133503; NM_133504; NM_133505; NM_133506; NM_133507	decorin isoform a preproprotein; decorin isoform b precursor; decorin isoform c precursor; decorin isoform d precursor; decorin isoform e precursor	organogenesis	0.1630	6.14	0.0035
NM_004718	NM_004718	cytochrome c oxidase subunit VIIa polypeptide 2 like	electron transport	0.1635	6.12	0.0031
AF130110		PRO1633		0.1636	6.11	0.0053
NM_001469	NM_001469	thyroid autoantigen 70kDa (Ku antigen)	DNA ligation; double-strand break repair via nonhomologous end-joining	0.1638	6.10	0.0062
NM_031916	NM_031916	AKAP-associated sperm protein	signal transduction	0.1638	6.10	0.0023
NM_000462	NM_000462; NM_130838; NM_130839	ubiquitin protein ligase E3A isoform 2; ubiquitin protein ligase E3A isoform 1; ubiquitin protein ligase E3A isoform 3	brain development; proteolysis and peptidolysis; ubiquitin-dependent protein catabolism	0.1640	6.10	0.0014
NM_003380	NM_003380	vimentin		0.1640	6.10	0.0075
NM_001338	NM_001338	coxsackie virus and adenovirus receptor		0.1641	6.09	0.0049
NM_015224	NM_015224	retinoblastoma-associated protein 140		0.1644	6.08	0.0182
NM_006530	NM_006530	glioma-amplified sequence-41	regulation of transcription, DNA-dependent	0.1645	6.08	0.0031
AK057940	NM_152524	tripin		0.1646	6.07	0.0068
NM_013264	NM_013264	DEAD (Asp-Glu-Ala-Asp) box polypeptide 25	regulation of translation; spermatogenesis	0.1653	6.05	0.0101
NM_031915	NM_031915	CLLL8 protein	chromatin modification	0.1653	6.05	0.0034
NM_004279	NM_004279	peptidase (mitochondrial processing) beta	proteolysis and peptidolysis	0.1655	6.04	0.0076
NM_006659	NM_006659	tubulin, gamma complex associated protein 2	microtubule nucleation; protein complex assembly	0.1657	6.03	0.0112
AB014731	NM_003677	density-regulated protein	cell growth and/or maintenance; translational initiation; biological_process unknown	0.1657	6.03	0.0001
NM_052811	NM_005798; NM_052811	ret finger protein 2	morphogenesis; negative regulation of cell cycle	0.1657	6.03	0.0007
NM_012260	NM_012260	2-hydroxyphytanoyl-CoA lyase	lipid metabolism	0.1660	6.02	0.0152
NM_000981	NM_000981	ribosomal protein L19	protein biosynthesis	0.1662	6.02	0.0177
NM_002887	NM_002887	arginyl-tRNA synthetase	arginyl-tRNA aminoacylation	0.1664	6.01	0.0030
NM_015646	NM_015646	RAP1B, member of RAS oncogene family	small GTPase mediated signal transduction	0.1665	6.01	0.0034
NM_001025	NM_001025	ribosomal protein S23	protein biosynthesis	0.1669	5.99	0.0057
AB058696	NM_022742	hypothetical protein DKFZp434G156		0.1671	5.98	0.0072
NM_003069	NM_003069; NM_139035	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a1 isoform a; SWI/SNF-related matrix-associated actin-dependent regulator of chromatin a1 isoform b	chromatin remodeling; regulation of transcription, DNA-dependent	0.1674	5.97	0.0021
NM_001568	NM_001568	murine mammary tumor integration site 6 (oncogene homolog)	cell growth and/or maintenance; regulation of translational initiation; protein biosynthesis	0.1676	5.97	0.0022

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AK055378				0.1685	5.93	0.0051
NM_016042	NM_016042	exosome component Rrp40	rRNA processing	0.1695	5.90	0.0040
AK027339	NM_022118	cutaneous T-cell lymphoma tumor antigen se70-2		0.1695	5.90	0.0051
NM_004365	NM_004365	centrin 3	centrosome cycle; mitosis; cytokinesis	0.1696	5.90	0.0079
AB023420	NM_002154; NM_198431	heat shock 70kDa protein 4 isoform a; heat shock 70kDa protein 4 isoform b		0.1696	5.89	0.0016
NM_005917	NM_005917	cytosolic malate dehydrogenase	tricarboxylic acid cycle	0.1697	5.89	0.0006
NM_005030	NM_005030	polo-like kinase	mitosis; protein amino acid phosphorylation; regulation of cell cycle	0.1698	5.89	0.0045
NM_003022	NM_003022	SH3 domain binding glutamic acid-rich protein like		0.1703	5.87	0.0002
NM_024410	NM_024410	outer dense fiber of sperm tails 1	spermatogenesis	0.1706	5.86	0.0080
NM_006807	NM_006807	chromobox homolog 1 (HP1 beta homolog *Drosophila*)	chromatin assembly/disassembly	0.1709	5.85	0.0002
NM_006471	NM_006471	myosin regulatory light chain MRCL3	muscle development; regulation of smooth muscle contraction	0.1711	5.85	0.0080
NM_014841	NM_014841	synaptosomal-associated protein, 91kDa homolog		0.1713	5.84	0.0054
AL136640	NM_016480	poly(A) binding protein interacting protein 2		0.1719	5.82	0.0061
M90355		BTF3 homologue	regulation of transcription, DNA-dependent	0.1722	5.81	0.0064
NM_019081	NM_014647; NM_019081	limkain b1 isoform 1; limkain b1 isoform 2		0.1724	5.80	0.0004
NM_002890	NM_002890; NM_022650	RAS p21 protein activator 1 isoform 1; RAS p21 protein activator 1 isoform 2	cell growth and/or maintenance; intracellular signaling cascade	0.1728	5.79	0.0001
NM_002615	NM_002615	serine (or cysteine) proteinase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1	development; neurogenesis; cell proliferation; negative regulation of angiogenesis	0.1732	5.77	0.0014
NM_022333	NM_003252; NM_022333	TIA1 cytotoxic granule-associated RNA-binding protein-like 1 isoform 1; TIA1 cytotoxic granule-associated RNA-binding protein-like 1 isoform 2	induction of apoptosis; regulation of transcription from Pol II promoter	0.1736	5.76	0.0020
BC017585	NM_178861	zinc finger protein 183-like 1		0.1738	5.75	0.0086
NM_018261	NM_018261; NM_178237	Sec3-like isoform 1; Sec3-like isoform 2	intracellular protein transport; exocytosis	0.1741	5.74	0.0003
AK023009	NM_152437	hypothetical protein DKFZp761B128		0.1741	5.74	0.0007
NM_004859	NM_004859	clathrin heavy chain	intracellular protein transport	0.1742	5.74	0.0023
NM_004853	NM_004853	syntaxin 8	transport; nonselective vesicle transport	0.1746	5.73	0.0070
NM_025180	NM_025180	hypothetical protein FLJ13386		0.1749	5.72	0.0146
NM_023934	NM_023934	hepatitis C virus core-binding protein 6		0.1750	5.71	0.0030
NM_000108	NM_000108	dihydrolipoamide dehydrogenase precursor	electron transport; glycolysis; energy pathways	0.1750	5.71	0.0014
NM_002803	NM_002803	proteasome 26S ATPase subunit 2	proteolysis and peptidolysis	0.1752	5.71	0.0128
NM_003653	NM_003653	COP9 constitutive photomorphogenic homolog subunit 3	response to light; signal transduction	0.1758	5.69	0.0005
NM_017948	NM_017948	chromosome 9 open reading frame 34		0.1759	5.68	0.0052
AK055130	NM_001743	calmodulin 2		0.1759	5.68	0.0101
NM_002954	NM_002954	ubiquitin and ribosomal protein S27a precursor	protein biosynthesis; biological_process unknown	0.1760	5.68	0.0002

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_005174	NM_005174	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	proton transport; energy pathways; ATP biosynthesis	0.1763	5.67	0.0031
AL049987				0.1766	5.66	0.0062
AF183421	NM_006868	RAB31, member RAS oncogene family	protein transport; small GTPase mediated signal transduction	0.1776	5.63	0.0033
NM_031301	NM_031301	presenilin stabilization factor-like		0.1779	5.62	0.0052
BC015858		ZNF37A protein	regulation of transcription, DNA-dependent	0.1786	5.60	0.0116
NM_001151	NM_001151	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	mitochondrial genome maintenance; small molecule transport; energy pathways	0.1786	5.60	0.0046
NM_000534	NM_000534	postmeiotic segregation 1	mismatch repair; oncogenesis	0.1789	5.59	0.0024
NM_019073	NM_019073	spermatogenesis associated 6		0.1793	5.58	0.0030
NM_000993	NM_000993	ribosomal protein L31	protein biosynthesis	0.1793	5.58	0.0151
NM_001029	NM_001029	ribosomal protein S26	regulation of transcription, DNA-dependent; protein biosynthesis	0.1800	5.56	0.0012
NM_016167	NM_016167	retinoic acid repressible protein		0.1801	5.55	0.0061
AB037801	NM_004241	thyroid hormone receptor interactor 8	regulation of transcription, DNA-dependent	0.1802	5.55	0.0018
NM_018448	NM_018448	TIP120 protein		0.1807	5.53	0.0063
NM_003375	NM_003375	voltage-dependent anion channel 2	anion transport	0.1807	5.53	0.0005
NM_015415	NM_015415	DKFZP564B167 protein		0.1810	5.53	0.0007
AL136755	NM_032132	hypothetical protein DKFZp434A1315		0.1819	5.50	0.0016
NM_015355	NM_015355	joined to JAZF1	cell growth and/or maintenance; regulation of transcription, DNA-dependent; chromatin modification	0.1820	5.49	0.0025
NM_024835	NM_024835	C3HC4-type zinc finger protein		0.1820	5.49	0.0071
NM_016047	NM_016047	pre-mRNA branch site protein p14	nuclear mRNA splicing, via spliceosome	0.1821	5.49	0.0037
NM_003816	NM_003816	a disintegrin and metalloproteinase domain 9 preproprotein	protein kinase cascade; proteolysis and peptidolysis	0.1826	5.48	0.0117
NM_002482	NM_002482; NM_152298; NM_172164	nuclear autoantigenic sperm protein isoform 2; nuclear autoantigenic sperm protein isoform 3; nuclear autoantigenic sperm protein isoform 1	DNA packaging; spermatogenesis	0.1826	5.48	0.0132
NM_002762	NM_002762	protamine 2	mitotic chromosome condensation; chromosome organization and biogenesis (sensu Eukarya); DNA packaging; spermatogenesis	0.1827	5.47	0.0101
NM_001154	NM_001154	annexin 5	blood coagulation	0.1832	5.46	0.0020
AK023362				0.1834	5.45	0.0067
NM_001277	NM_001277	choline kinase	lipid transport; lipid metabolism	0.1834	5.45	0.0058
NM_005410	NM_005410	selenoprotein P precursor	response to oxidative stress	0.1836	5.45	0.0058
AB037851		KIAA1430 protein		0.1840	5.43	0.0035
M62896				0.1840	5.43	0.0110
AK056690	NM_177452	trafficking protein particle complex 6B		0.1842	5.43	0.0063
NM_016422	NM_016422	ring finger protein 141		0.1843	5.43	0.0091
NM_005561	NM_005561	lysosomal-associated membrane protein 1		0.1848	5.41	0.0023
NM_005717	NM_005717	actin related protein 2/3 complex subunit 5	cell motility; actin cytoskeleton organization and biogenesis	0.1851	5.40	0.0019

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AF078164	NM_033276	Ku70-binding protein 3		0.1852	5.40	0.0002
NM_004234	NM_004234	zinc finger protein 93 homolog	regulation of transcription, DNA-dependent	0.1854	5.40	0.0051
NM_005499	NM_005499	SUMO-1 activating enzyme subunit 2	ubiquitin cycle	0.1855	5.39	0.0012
AF070559	NM_138779	hypothetical protein BC015148		0.1855	5.39	0.0172
NM_014970	NM_014970	kinesin-associated protein 3	signal transduction; protein complex assembly	0.1856	5.39	0.0186
NM_005869	NM_005869	serologically defined colon cancer antigen 10	protein folding	0.1857	5.38	0.0106
NM_005731	NM_005731; NM_152862	actin related protein 2/3 complex subunit 2	cell motility	0.1858	5.38	0.0143
NM_004300	NM_004300; NM_007099; NM_177554	acid phosphatase 1 isoform c; acid phosphatase 1 isoform b; acid phosphatase 1 isoform a		0.1859	5.38	0.0020
NM_014758	NM_014758	sorting nexin 19	intracellular protein transport; intracellular signaling cascade	0.1863	5.37	0.0058
NM_002092	NM_002092	G-rich RNA sequence binding factor 1	mRNA polyadenylation	0.1864	5.36	0.0012
NM_001494	NM_001494	GDP dissociation inhibitor 2	protein transport; nonselective vesicle transport	0.1865	5.36	0.0051
BC017197	NM_021960; NM_182763	myeloid cell leukemia sequence 1 isoform 1; myeloid cell leukemia sequence 1 isoform 2	apoptotic program; heat shock response; development	0.1867	5.36	0.0002
NM_016302	NM_016302	protein x 0001	ATP-dependent proteolysis	0.1869	5.35	0.0028
NM_020122	NM_020122	potassium channel modulatory factor 1		0.1869	5.35	0.0048
D29641	NM_015360	KIAA0052 protein		0.1872	5.34	0.0034
NM_006428	NM_006428	mitochondrial ribosomal protein L28		0.1875	5.33	0.0067
NM_018359	NM_018359	hypothetical protein FLJ11200		0.1876	5.33	0.0002
NM_017819	NM_017819	hypothetical protein FLJ20432		0.1877	5.33	0.0026
NM_015344	NM_015344	leptin receptor overlapping transcript-like 1		0.1881	5.32	0.0081
NM_001009	NM_001009	ribosomal protein S5	protein biosynthesis	0.1881	5.32	0.0076
AK055864	NM_018418	spermatogenesis-associated protein 7		0.1883	5.31	0.0136
NM_032461	NM_032461	sperm protein associated with the nucleus, X chromosome, family member B1		0.1887	5.30	0.0022
AK023044	NM_016290	retinoid x receptor interacting protein		0.1888	5.30	0.0071
NM_021076	NM_021076	neurofilament, heavy polypeptide 200kDa		0.1889	5.29	0.0173
NM_002715	NM_002715	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	regulation of cell cycle; protein amino acid dephosphorylation	0.1895	5.28	0.0051
NM_007027	NM_007027	topoisomerase (DNA) II binding protein	DNA replication and chromosome cycle; DNA metabolism	0.1897	5.27	0.0047
NM_001697	NM_001697	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)	proton transport; ATP biosynthesis	0.1898	5.27	0.0020
AK056390	NM_018846	SBBI26 protein		0.1899	5.27	0.0000
NM_014463	NM_014463	Lsm3 protein	nuclear mRNA splicing, via spliceosome	0.1900	5.26	0.0040
AB040938		KIAA1505 protein		0.1901	5.26	0.0086
NM_016589	NM_016589	M5-14 protein		0.1901	5.26	0.0030
NM_007214	NM_007214	SEC63-like protein	protein-membrane targeting	0.1904	5.25	0.0032
NM_025165	NM_025165	elongation factor RNA polymerase II-like 3		0.1904	5.25	0.0060
AF333334	NM_005893	calicin	spermatogenesis	0.1904	5.25	0.0030

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_013291	NM_013291	cleavage and polyadenylation specific factor 1, 160kDa	mRNA polyadenylation; mRNA cleavage	0.1905	5.25	0.0065
NM_003914	NM_003914	cyclin A1	male meiosis I; regulation of CDK activity; mitosis; spermatogenesis; cytokinesis	0.1907	5.24	0.0124
NM_006988	NM_006988	a disintegrin and metalloprotease with thrombospondin motifs-1 preproprotein	integrin-mediated signaling pathway; negative regulation of cell proliferation; proteolysis and peptidolysis	0.1909	5.24	0.0075
NM_016038	NM_016038	Shwachman-Bodian-Diamond syndrome protein		0.1912	5.23	0.0059
AK058131	NM_152567	hypothetical protein FLJ25402		0.1918	5.21	0.0032
NM_016013	NM_016013	CGI-65 protein		0.1920	5.21	0.0055
AB059277	NM_015446	transcription factor ELYS		0.1921	5.20	0.0015
AL137385				0.1922	5.20	0.0016
NM_003187	NM_003187; NM_016283	TBP-associated factor 9; adrenal gland protein AD-004	transcription from Pol II promoter	0.1922	5.20	0.0105
NM_014670	NM_014670	basic leucine zipper and W2 domains 1		0.1923	5.20	0.0044
NM_003753	NM_003753	eukaryotic translation initiation factor 3 subunit 7	regulation of translational initiation; protein biosynthesis	0.1926	5.19	0.0083
NM_005100	NM_005100; NM_144497	A-kinase anchor protein 12 isoform 1; A-kinase anchor protein 12 isoform 2	G-protein linked receptor protein signalling pathway	0.1927	5.19	0.0053
NM_001444	NM_001444	fatty acid binding protein 5 (psoriasis-associated)	epidermal differentiation; transport; lipid metabolism	0.1932	5.18	0.0004
NM_014781	NM_014781	Rb1-inducible coiled coil protein 1		0.1932	5.18	0.0088
NM_013232	NM_013232	programmed cell death 6	induction of apoptosis by extracellular signals	0.1933	5.17	0.0042
NM_004728	NM_004728	DEAD (Asp-Glu-Ala-Asp) box polypeptide 21		0.1933	5.17	0.0033
NM_014222	NM_014222	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa		0.1934	5.17	0.0017
AB067512		KIAA1925 protein	DNA recombination	0.1935	5.17	0.0030
NM_017613	NM_017613; NM_145794; NM_145795	downstream neighbor of SON isoform a; downstream neighbor of SON isoform b; downstream neighbor of SON isoform c		0.1938	5.16	0.0002
AK026479	NM_020468; NM_153816	sorting nexin 14 isoform b; sorting nexin 14 isoform a	intracellular protein transport; intracellular signaling cascade	0.1938	5.16	0.0314
NM_005744	NM_005744	ariadne homolog, ubiquitin-conjugating enzyme E2 binding protein, 1	ubiquitin-dependent protein catabolism	0.1940	5.15	0.0012
NM_016227	NM_014283; NM_016227	chromosome 1 open reading frame 9; membrane protein CH1		0.1940	5.15	0.0001
AK057504	NM_144594	hypothetical protein FLJ32942		0.1941	5.15	0.0075
NM_003798	NM_003798	catenin (cadherin-associated protein), alpha-like 1	apoptosis	0.1943	5.15	0.0041
NM_002079	NM_002079	aspartate aminotransferase 1	aspartate catabolism; amino acid metabolism; biosynthesis	0.1947	5.14	0.0047
AK022565	NM_138423; NM_177974	H63 breast cancer expressed gene isoform a; H63 breast cancer expressed gene isoform b		0.1948	5.13	0.0070
AB046781	NM_018003	uveal autoantigen with coiled-coil domains and ankyrin repeats	biological_process unknown	0.1949	5.13	0.0025
NM_006422	NM_006422	A-kinase anchor protein 3	acrosome reaction; cell motility	0.1954	5.12	0.0178
NM_000900	NM_000900	matrix Gla protein	ossification; cartilage condensation	0.1954	5.12	0.0116
NM_014160	NM_014160	makorin, ring finger protein, 2	biological_process unknown	0.1955	5.12	0.0058

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_018694	NM_016638; NM_018694	SRp25 nuclear protein	RNA splicing	0.1956	5.11	0.0210
NM_002810	NM_002810; NM_153822	proteasome 26S non-ATPase subunit 4 isoform 1; proteasome 26S non-ATPase subunit 4 isoform 2		0.1958	5.11	0.0155
AK027572	NM_153331	potassium channel tetramerisation domain containing 6	potassium ion transport	0.1960	5.10	0.0045
NM_000291	NM_000291	phosphoglycerate kinase 1		0.1960	5.10	0.0071
NM_002805	NM_002805	proteasome 26S ATPase subunit 5	signal transduction; transcription from Pol II promoter; protein catabolism	0.1961	5.10	0.0078
NM_001239	NM_001239	cyclin H	regulation of CDK activity; cell cycle; regulation of transcription, DNA-dependent; DNA repair	0.1961	5.10	0.0011
NM_016053	NM_016053	CGI-116 protein		0.1961	5.10	0.0017
NM_002748	NM_002748	mitogen-activated protein kinase 6	cell cycle; protein amino acid phosphorylation; signal transduction	0.1965	5.09	0.0010
NM_000194	NM_000194	hypoxanthine phosphoribosyltransferase 1	purine salvage; nucleoside metabolism; behavior	0.1967	5.08	0.0018
NM_000290	NM_000290	phosphoglycerate mutase 2 (muscle)	striated muscle contraction; glycolysis; metabolism	0.1970	5.08	0.0004
Y14039	NM_003879	CASP8 and FADD-like apoptosis regulator	induction of apoptosis by extracellular signals; anti-apoptosis; proteolysis and peptidolysis	0.1971	5.07	0.0024
BC010282	NM_133259	leucine-rich PPR motif-containing protein	biological_process unknown	0.1972	5.07	0.0002
AF272348		zinc-finger protein DZIPt1		0.1973	5.07	0.0085
AB029316	NM_015435; NM_183419	ring finger protein 19	microtubule cytoskeleton organization and biogenesis; protein modification	0.1975	5.06	0.0017
D80010	NM_145693	lipin 1	adipocyte differentiation	0.1977	5.06	0.0002
NM_000904	NM_000904	NAD(P)H dehydrogenase, quinone 2	electron transport	0.1983	5.04	0.0074
AK000872				0.1984	5.04	0.0013
NM_002970	NM_002970	spermidine/spermine N1-acetyltransferase		0.1988	5.03	0.0004
NM_021103	NM_021103	thymosin, beta 10	spermatid development; cytoskeleton organization and biogenesis	0.1992	5.02	0.0091
AF226044	NM_017719	HSNFRK		0.1992	5.02	0.0021
AB023198	NM_015040	phosphatidylinositol-3-phosphate/phosphatidylinositol 5-kinase, type III	intracellular signaling cascade	0.1994	5.02	0.0009
NM_033124	NM_033124	NYD-SP28 protein		0.1998	5.00	0.0088
NM_031943	NM_031943	IFP38	translational initiation	0.1999	5.00	0.0059
NM_005313	NM_005313	glucose regulated protein, 58kDa	protein-nucleus import; protein-ER retention; electron transport; signal transduction	0.2001	5.00	0.0025
NM_002902	NM_002902	reticulocalbin 2, EF-hand calcium binding domain		0.2003	4.99	0.0050
AB046821	NM_007159	sarcolemma associated protein	muscle contraction	0.2004	4.99	0.0019
AK058139	NM_144605	hypothetical protein FLJ25410	cell cycle	0.2004	4.99	0.0121
NM_003137	NM_003137	SFRS protein kinase 1	protein amino acid phosphorylation	0.2004	4.99	0.0008
AY027808	NM_130784	hypothetical telomeric protein		0.2005	4.99	0.0074
NM_016447	NM_016447	membrane protein, palmitoylated 6	protein complex assembly	0.2005	4.99	0.0000
NM_006002	NM_006002	ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)	ubiquitin-dependent protein catabolism	0.2006	4.99	0.0000

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_015251	NM_015251	KIAA0431 protein		0.2008	4.98	0.0035
NM_006870	NM_006870	destrin (actin depolymerizing factor)	actin polymerization and/or depolymerization	0.2008	4.98	0.0084
NM_002056	NM_002056	glucosamine-fructose-6-phosphate aminotransferase	energy reserve metabolism; fructose 6-phosphate metabolism; carbohydrate metabolism; glutamine metabolism	0.2010	4.97	0.0019
NM_003948	NM_003948	cyclin-dependent kinase-like 2	sex differentiation; protein amino acid phosphorylation; signal transduction	0.2011	4.97	0.0021
NM_018844	NM_018844	B-cell receptor-associated protein BAP29	intracellular protein transport; apoptosis	0.2012	4.97	0.0051
NM_006585	NM_006585	chaperonin containing TCP1, subunit 8 (theta)		0.2013	4.97	0.0018
NM_003796	NM_003796; NM_134447	RPB5-mediating protein isoform a; RPB5-mediating protein isoform b	pathogenesis; protein folding; regulation of transcription from Pol II promoter	0.2014	4.97	0.0084
NM_013233	NM_013233	serine threonine kinase 39 (STE20/SPS1 homolog, yeast)	response to stress; protein amino acid phosphorylation	0.2015	4.96	0.0028
NM_006417	NM_006417	interferon-induced, hepatitis C-associated microtubular aggregat	response to virus; invasive growth	0.2015	4.96	0.0053
NM_006294	NM_006294	ubiquinol-cytochrome c reductase binding protein	oxidative phosphorylation; aerobic respiration; electron transport; mitochondrial electron transport, ubiquinol to cytochrome c	0.2017	4.96	0.0136
AK026843				0.2017	4.96	0.0003
NM_016607	NM_016607; NM_177947; NM_177948	ALEX3 protein		0.2017	4.96	0.0026
NM_032368	NM_032368	leucine zipper and CTNNBIP1 domain containing		0.2018	4.95	0.0053
AF039442		colon cancer antigen NY-CO-45		0.2020	4.95	0.0054
NM_004708	NM_004708	programmed cell death 5	induction of apoptosis	0.2021	4.95	0.0078
NM_004446	NM_004446	glutamyl-prolyl tRNA synthetase	glutamyl-tRNA aminoacylation; prolyl-tRNA aminoacylation; protein complex assembly	0.2022	4.95	0.0094
NM_003118	NM_003118	secreted protein, acidic, cysteine-rich (osteonectin)	ossification	0.2026	4.94	0.0264
NM_018322	NM_018322	hypothetical protein FLJ11101		0.2026	4.94	0.0105
NM_002262	NM_002262; NM_007334	killer cell lectin-like receptor subfamily D, member 1 isoform 1; killer cell lectin-like receptor subfamily D, member 1 isoform 2	cell surface receptor linked signal transduction; heterophilic cell adhesion; antimicrobial humoral response (sensu Vertebrata)	0.2027	4.93	0.0003
NM_015895	NM_015895	geminin	negative regulation of DNA replication; cell cycle arrest	0.2030	4.93	0.0019
NM_014372	NM_014372	ring finger protein 11	oncogenesis	0.2032	4.92	0.0006
NM_007054	NM_007054	kinesin family member 3A	organelle organization and biogenesis	0.2033	4.92	0.0000
NM_002355	NM_002355	cation-dependent mannose-6-phosphate receptor precursor	endosome to lysosome transport; receptor mediated endocytosis	0.2034	4.92	0.0015
NM_004339	NM_004339	pituitary tumor-transforming gene 1 protein-interacting protein precursor	protein-nucleus import; development	0.2034	4.92	0.0077
NM_003462	NM_003462	axonemal dynein light chain	cell motility; fertilization	0.2035	4.91	0.0049

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AK057157	NM_002869; NM_198896	RAB6A, member RAS oncogene family isoform a; RAB6A, member RAS oncogene family isoform b	small GTPase mediated signal transduction; nonselective vesicle transport; intracellular protein transport	0.2036	4.91	0.0026
AF218029	NM_005324	H3 histone, family 3B	nucleosome assembly; chromosome organization and biogenesis (sensu Eukarya)	0.2037	4.91	0.0097
NM_001030	NM_001030	ribosomal protein S27	protein biosynthesis; cell proliferation; signal transduction	0.2038	4.91	0.0095
NM_003908	NM_003908	eukaryotic translation initiation factor 2 beta	translational initiation	0.2044	4.89	0.0029
NM_000179	NM_000179	mutS homolog 6	oncogenesis; mismatch repair	0.2048	4.88	0.0027
NM_005752	NM_005752	C-type lectin, superfamily member 1	skeletal development; heterophilic cell adhesion	0.2049	4.88	0.0036
AK027789	NM_178862	source of immunodominant MHC-associated peptides	protein amino acid glycosylation	0.2054	4.87	0.0000
NM_007045	NM_007045; NM_194429	FGFR1 oncogene partner isoform a; FGFR1 oncogene partner isoform b	positive regulation of cell proliferation; oncogenesis	0.2054	4.87	0.0076
NM_001028	NM_001028	ribosomal protein S25	protein biosynthesis	0.2054	4.87	0.0213
NM_001068	NM_001068	DNA topoisomerase II, beta isozyme	regulation of transcription, DNA-dependent; DNA metabolism; DNA topological change	0.2054	4.87	0.0029
NM_004280	NM_004280	eukaryotic translation elongation factor 1 epsilon 1	protein biosynthesis	0.2059	4.86	0.0144
NM_002266	NM_002266	karyopherin alpha 2	regulation of DNA recombination; G2 phase of mitotic cell cycle; NLS-bearing substrate-nucleus import; intracellular protein transport; DNA metabolism; M-phase specific microtubule process	0.2060	4.86	0.0180
NM_018365	NM_018365	meiosis-specific nuclear structural protein 1		0.2061	4.85	0.0083
NM_014269	NM_014269; NM_021779; NM_021780	a disintegrin and metalloproteinase domain 29 isoform 1 preproprotein; a disintegrin and metalloproteinase domain 29 isoform 3 preproprotein; a disintegrin and metalloproteinase domain 29 isoform 2 preproprotein	proteolysis and peptidolysis; spermatogenesis	0.2062	4.85	0.0071
NM_022551	NM_022551	ribosomal protein S18	protein biosynthesis	0.2062	4.85	0.0149
NM_002904	NM_002904	RD RNA-binding protein		0.2067	4.84	0.0056
AK001693	NM_015420	DKFZP564O0463 protein		0.2071	4.83	0.0054
NM_032597	NM_032597	testes development-related NYD-SP21		0.2072	4.83	0.0030
NM_032598	NM_032598	testes development-related NYD-SP20		0.2072	4.83	0.0098
NM_023008	NM_023008; NM_178159	hypothetical protein FLJ12949 isoform 1; hypothetical protein FLJ12949 isoform 2		0.2075	4.82	0.0040
NM_004396	NM_004396	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	cell growth	0.2075	4.82	0.0177
AB011173	NM_015013	KIAA0601 protein		0.2078	4.81	0.0081
AL050018	NM_015525	hypothetical protein		0.2079	4.81	0.0074
NM_001539	NM_001539	DnaJ (Hsp40) homolog, subfamily A, member 1	protein folding; heat shock response	0.2083	4.80	0.0088
NM_005745	NM_005745	B-cell receptor-associated protein 31	intracellular protein transport; apoptosis; immune response	0.2086	4.79	0.0113
NM_001122	NM_001122	adipose differentiation-related protein		0.2086	4.79	0.0278

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_015909	NM_015909	neuroblastoma-amplified protein	protein biosynthesis	0.2087	4.79	0.0103
AL136719	NM_006717	spindlin	gametogenesis	0.2087	4.79	0.0013
NM_015653	NM_015653	hypothetical protein MGC4107		0.2088	4.79	0.0031
NM_078469	NM_016567; NM_078468; NM_078469	BRCA2 and CDKN1A-interacting protein isoform BCCIPalpha; BRCA2 and CDKN1A-interacting protein isoform BCCIPbeta; BRCA2 and CDKN1A-interacting protein isoform C	regulation of CDK activity	0.2089	4.79	0.0190
NM_031446	NM_031446	hypothetical protein PNAS-131		0.2089	4.79	0.0130
BC001082	NM_015448	DKFZP566F084 protein		0.2089	4.79	0.0009
BC011981	NM_138414	hypothetical protein BC011981		0.2090	4.78	0.0113
AY009106	NM_015668	DKFZP434I092 protein		0.2092	4.78	0.0133
AK055491	NM_001378	dynein, cytoplasmic, intermediate polypeptide 2	microtubule-based movement	0.2099	4.77	0.0098
NM_021634	NM_021634	leucine-rich repeat-containing G protein-coupled receptor 7	G-protein coupled receptor protein signaling pathway; signal transduction	0.2099	4.76	0.0020
NM_016271	NM_016271; NM_198128	ring finger protein 138 isoform 1; ring finger protein 138 isoform 2		0.2100	4.76	0.0142
NM_031210	NM_031210	hypothetical protein DC50		0.2101	4.76	0.0084
NM_018186	NM_018186	hypothetical protein FLJ10706	proteolysis and peptidolysis	0.2103	4.76	0.0007
NM_018693	NM_012167; NM_018693; NM_025133	F-box only protein 11 isoform 3; F-box only protein 11 isoform 2; F-box only protein 11 isoform 1	ubiquitin cycle; ubiquitin-dependent protein catabolism; protein modification	0.2103	4.76	0.0018
NM_003715	NM_003715	vesicle docking protein p115	nonselective vesicle docking; intracellular protein transport	0.2109	4.74	0.0089
AL132665	NM_004331	BCL2/adenovirus E1B 19kD-interacting protein 3-like	negative regulation of survival gene products; induction of apoptosis	0.2116	4.73	0.0033
NM_004261	NM_004261	15 kDa selenoprotein		0.2118	4.72	0.0023
NM_005006	NM_005006	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa precursor	complex I (NADH to ubiquinone)	0.2119	4.72	0.0017
NM_000365	NM_000365	triosephosphate isomerase 1	pentose-phosphate shunt; glycolysis; gluconeogenesis; metabolism; fatty acid biosynthesis	0.2119	4.72	0.0129
AK058124	NM_144722	hypothetical protein FLJ25395		0.2119	4.72	0.0047
AL137345				0.2125	4.71	0.0076
NM_012341	NM_012341	G protein-binding protein CRFG		0.2127	4.70	0.0078
NM_003618	NM_003618	mitogen-activated protein kinase kinase kinase kinase 3	protein kinase cascade; response to stress; protein amino acid phosphorylation	0.2127	4.70	0.0008
NM_015904	NM_015904	translation initiation factor IF2	regulation of transcription, DNA-dependent; regulation of translational initiation; protein biosynthesis	0.2129	4.70	0.0059
NM_004766	NM_004766	coatomer protein complex, subunit beta 2 (beta prime)	intracellular protein transport; exocytosis	0.2129	4.70	0.0004
NM_020186	NM_020186	ACN9 homolog		0.2132	4.69	0.0152
NM_021114	NM_021114	serine protease inhibitor, Kazal type, 2 (acrosin-trypsin inhibitor)		0.2133	4.69	0.0126
NM_000018	NM_000018	acyl-Coenzyme A dehydrogenase, very long chain precursor	fatty acid beta-oxidation; energy derivation by oxidation of organic compounds; electron transport; fatty acid metabolism	0.2134	4.69	0.0065
NM_002951	NM_002951	ribophorin II precursor	protein modification	0.2135	4.68	0.0021
AB037810		KIAA1389 protein		0.2137	4.68	0.0001

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_000988	NM_000988	ribosomal protein L27		0.2139	4.68	0.0093
NM_006696	NM_006696; NM_139199; NM_183359	bromodomain containing 8 isoform 1; bromodomain containing 8 isoform 2; bromodomain containing 8 isoform 3	cell surface receptor linked signal transduction; regulation of transcription from Pol II promoter	0.2139	4.67	0.0022
AL162039				0.2142	4.67	0.0009
NM_003755	NM_003755	eukaryotic translation initiation factor 3, subunit 4 delta, 44kDa	regulation of translational initiation; protein biosynthesis	0.2146	4.66	0.0079
NM_032891	NM_032891	WD domain, G-beta repeat-containing protein		0.2148	4.66	0.0061
NM_006429	NM_006429	chaperonin containing TCP1, subunit 7 (eta)	protein folding; regulation of cell cycle	0.2148	4.65	0.0107
AK057820				0.2150	4.65	0.0103
AJ306929	NM_024524	ATPase family homolog up-regulated in senescence cells	cation transport; metabolism	0.2150	4.65	0.0017
BC010062		Unknown (protein for IMAGE:3953648)		0.2152	4.65	0.0119
NM_016127	NM_016127	hypothetical protein MGC8721		0.2152	4.65	0.0131
AL079314	NM_014455	zinc finger protein 364		0.2153	4.64	0.0079
NM_024098	NM_024098	hypothetical protein MGC2574	regulation of transcription, DNA-dependent	0.2156	4.64	0.0114
NM_032364	NM_032364	dopamine receptor interacting protein		0.2157	4.64	0.0007
NM_004168	NM_004168	succinate dehydrogenase complex, subunit A, flavoprotein precursor	aerobic respiration; tricarboxylic acid cycle	0.2158	4.63	0.0069
NM_016106	NM_016106; NM_182835	vesicle transport-related protein isoform a; vesicle transport-related protein isoform b		0.2160	4.63	0.0028
NM_002013	NM_002013	FK506-binding protein 3	protein folding	0.2163	4.62	0.0133
AF353674	NM_033271	BTB domain protein BDPL		0.2168	4.61	0.0047
NM_024606	NM_024606	hypothetical protein FLJ11756		0.2169	4.61	0.0000
NM_032288	NM_032288	hypothetical protein DKFZp761B1514		0.2172	4.60	0.0029
NM_016587	NM_007276; NM_016587	chromobox homolog 3	chromatin assembly/disassembly; regulation of transcription, DNA-dependent; chromatin modification	0.2172	4.60	0.0014
NM_002792	NM_002792; NM_152255	proteasome alpha 7 subunit isoform 1; proteasome alpha 7 subunit isoform 2	ubiquitin-dependent protein catabolism	0.2175	4.60	0.0064
NM_031157	NM_002136; NM_031157	heterogeneous nuclear ribonucleoprotein A1 isoform a; heterogeneous nuclear ribonucleoprotein A1 isoform b	RNA-nucleus export; mRNA processing	0.2175	4.60	0.0155
NM_016141	NM_016141	dynein light chain-A	small GTPase mediated signal transduction	0.2176	4.60	0.0030
AK058070				0.2176	4.60	0.0017
AL354915				0.2176	4.60	0.0093
NM_007263	NM_007263	coatomer protein complex, subunit epsilon	nonselective vesicle transport; intracellular protein transport; antimicrobial humoral response (sensu Vertebrata)	0.2177	4.59	0.0023
AB037741	NM_020771	HECT domain and ankyrin repeat containing, E3 ubiquitin protein ligase 1	ubiquitin cycle	0.2180	4.59	0.0134
AL161957	NM_017934	pleckstrin homology domain interacting protein		0.2182	4.58	0.0064
NM_004865	NM_004865	TBP-like 1	transcription initiation from Pol II promoter; regulation of transcription, DNA-dependent	0.2182	4.58	0.0041
AK058144	NM_144708	hypothetical protein FLJ25415		0.2182	4.58	0.0075

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AL110274	NM_003888; NM_170696; NM_170697	aldehyde dehydrogenase 1A2 isoform 1; aldehyde dehydrogenase 1A2 isoform 2; aldehyde dehydrogenase 1A2 isoform 3	vitamin A metabolism; oncogenesis	0.2183	4.58	0.0005
NM_020188	NM_020188	DC13 protein		0.2184	4.58	0.0011
AF073310	NM_003749	insulin receptor substrate 2		0.2191	4.56	0.0018
NM_014966	NM_014966; NM_138614; NM_138615	DEAH (Asp-Glu-Ala-His) box polypeptide 30 isoform 2; DEAH (Asp-Glu-Ala-His) box polypeptide 30 isoform 3; DEAH (Asp-Glu-Ala-His) box polypeptide 30 isoform 1		0.2191	4.56	0.0106
BC002971	NM_012073	chaperonin containing TCP1, subunit 5 (epsilon)	protein folding	0.2192	4.56	0.0165
NM_058216	NM_002876; NM_058216; NM_058217	RAD51 homolog C isoform 2; RAD51 homolog C isoform 1; RAD51 homolog C isoform 3	DNA repair; DNA recombination	0.2192	4.56	0.0005
AB053313				0.2192	4.56	0.0066
NM_001827	NM_001827	CDC28 protein kinase 2	regulation of CDK activity; cell cycle; cytokinesis	0.2196	4.55	0.0017
NM_001515	NM_001515	general transcription factor IIH, polypeptide 2, 44kD subunit	regulation of transcription, DNA-dependent; DNA repair	0.2196	4.55	0.0034
BC007344		Similar to RIKEN cDNA 4833415E20 gene		0.2198	4.55	0.0099
NM_032830	NM_032830	cirhin		0.2204	4.54	0.0032
NM_006304	NM_006304	candidate for split hand/foot malformation type 1	biological_process unknown	0.2204	4.54	0.0158
AL049246				0.2205	4.54	0.0008
AL080111	NM_133494	NIMA (never in mitosis gene a)-related kinase 7	protein amino acid phosphorylation	0.2205	4.54	0.0012
AK024475		FLJ00068 protein		0.2207	4.53	0.0027
NM_031858	NM_005899; NM_031858; NM_031862	membrane component, chromosome 17, surface marker 2		0.2210	4.52	0.0137
NM_003619	NM_003619	neurotrypsin precursor	proteolysis and peptidolysis	0.2213	4.52	0.0069
NM_006324	NM_006324	craniofacial development protein 1		0.2213	4.52	0.0092
NM_018997	NM_018997; NM_031901	mitochondrial ribosomal protein S21		0.2219	4.51	0.0023
NM_013320	NM_013320	host cell factor 2	viral life cycle; transport; regulation of transcription from Pol II promoter	0.2220	4.50	0.0037
NM_001358	NM_001358	DEAH (Asp-Glu-Ala-His) box polypeptide 15	nuclear mRNA splicing, via spliceosome	0.2221	4.50	0.0002
AB029551	NM_012234	RING1 and YY1 binding protein	regulation of transcription, DNA-dependent	0.2221	4.50	0.0050
NM_021129	NM_021129	inorganic pyrophosphatase	phosphate metabolism	0.2222	4.50	0.0006
NM_018297	NM_018297	N-glycanase 1		0.2226	4.49	0.0080
AL133596				0.2228	4.49	0.0023
AF033199				0.2230	4.49	0.0066
AB067486	NM_052903	tubulin, gamma complex associated protein 5	microtubule nucleation	0.2230	4.48	0.0043
NM_001685	NM_001685	ATP synthase, H+ transporting, mitochondrial precursor	proton transport; energy pathways	0.2231	4.48	0.0060
AK001134				0.2231	4.48	0.0011
NM_001022	NM_001022	ribosomal protein S19	hemocyte development; protein biosynthesis	0.2235	4.47	0.0123
NM_016359	NM_016359; NM_018454	clone HQ0310 PRO0310p1; nucleolar and spindle associated protein 1		0.2237	4.47	0.0022
NM_018491	NM_018491	COBW-like protein		0.2238	4.47	0.0133
AK024327				0.2241	4.46	0.0043

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_032930	NM_032930	hypothetical protein MGC13040		0.2242	4.46	0.0012
NM_006206	NM_006206	platelet-derived growth factor receptor alpha precursor	transmembrane receptor protein tyrosine kinase signaling pathway; protein amino acid phosphorylation; cell surface receptor linked signal transduction; cell proliferation	0.2243	4.46	0.0045
NM_001912	NM_001912; NM_145918	cathepsin L preproprotein	proteolysis and peptidolysis	0.2244	4.46	0.0025
AB037791		KIAA1370 protein		0.2244	4.46	0.0002
NM_003418	NM_003418	zinc finger protein 9	cholesterol biosynthesis; regulation of transcription, DNA-dependent	0.2248	4.45	0.0101
NM_003453	NM_003453; NM_197968	zinc finger protein 198		0.2248	4.45	0.0040
AK000315	NM_017758	hypothetical protein FLJ20308		0.2251	4.44	0.0041
NM_004336	NM_004336	BUB1 budding uninhibited by benzimidazoles 1 homolog	mitotic spindle checkpoint; cell cycle; mitosis; protein amino acid phosphorylation	0.2255	4.44	0.0023
NM_014883	NM_014883	family with sequence similarity 13, member A1		0.2260	4.43	0.0032
NM_014368	NM_014368	LIM homeobox protein 6 isoform 1	brain development; regulation of transcription, DNA-dependent	0.2260	4.42	0.0059
AK057204	NM_152415	hypothetical protein FLJ32642		0.2260	4.42	0.0078
AF263613		membrane-associated calcium-independent phospholipase A2 gamma		0.2264	4.42	0.0013
AY040873		C21orf57 isoform A protein		0.2264	4.42	0.0039
NM_001551	NM_001551	immunoglobulin (CD79A) binding protein 1	response to biotic stimulus; B-cell activation; regulation of signal transduction	0.2265	4.42	0.0038
NM_006251	NM_006251	protein kinase, AMP-activated, alpha 1 catalytic subunit	protein amino acid phosphorylation; signal transduction	0.2266	4.41	0.0012
NM_001618	NM_001618	poly(ADP-ribosyl)transferase	protein amino acid ADP-ribosylation; cell growth and/or maintenance; DNA repair; transcription from Pol II promoter	0.2267	4.41	0.0010
AL121886				0.2275	4.40	0.0135
AK055070	NM_174909	hypothetical protein MGC23909		0.2280	4.39	0.0023
NM_006292	NM_006292	tumor susceptibility gene 101	ubiquitin cycle; intracellular protein transport; regulation of cell growth	0.2280	4.39	0.0020
NM_013417	NM_002161; NM_013417	isoleucine-tRNA synthetase	isoleucyl-tRNA aminoacylation	0.2284	4.38	0.0029
NM_002717	NM_002717	alpha isoform of regulatory subunit B55, protein phosphatase 2	signal transduction; protein amino acid dephosphorylation	0.2284	4.38	0.0012
NM_005445	NM_005445	chondroitin sulfate proteoglycan 6 (bamacan)	mitotic spindle assembly; sister chromatid cohesion; meiosis; chromosome segregation; transport; cell cycle; DNA repair; signal transduction	0.2285	4.38	0.0008
NM_002018	NM_002018	flightless I homolog	muscle contraction; development	0.2287	4.37	0.0075
NM_012394	NM_012394	prefoldin 2	protein folding	0.2290	4.37	0.0122
AL050006	NM_015522; NM_016008	dynein 2 light intermediate chain isoform 2; dynein 2 light intermediate chain isoform 1		0.2293	4.36	0.0052
NM_018447	NM_018447	30 kDa protein		0.2295	4.36	0.0051

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_032273	NM_032273	hypothetical protein DKFZp586C1924		0.2296	4.36	0.0003
AK057575	NM_002635; NM_005888	phosphate carrier precursor isoform 1b; phosphate carrier precursor isoform 1a	transport; energy pathways	0.2297	4.35	0.0030
NM_005805	NM_005805	26S proteasome-associated pad1 homolog	ubiquitin-dependent protein catabolism	0.2297	4.35	0.0143
AB023163	NM_015336	huntingtin interacting protein 14		0.2298	4.35	0.0036
NM_023011	NM_023011; NM_080687	UPF3 regulator of nonsense transcripts homolog A isoform hUpf3p; UPF3 regulator of nonsense transcripts homolog A isoform hUpf3pdelta		0.2303	4.34	0.0013
NM_024570	NM_024570	hypothetical protein FLJ11712		0.2304	4.34	0.0021
NM_018171	NM_018171	DIP13 beta		0.2310	4.33	0.0054
NM_032043	NM_032043	BRCA1 interacting protein C-terminal helicase 1	nucleotide-excision repair	0.2311	4.33	0.0063
BC001511	NM_007277	Sec6 protein	intracellular protein transport; exocytosis	0.2312	4.32	0.0027
AK054750				0.2316	4.32	0.0006
NM_002574	NM_002574; NM_181696; NM_181697	peroxiredoxin 1	skeletal development; cell proliferation	0.2317	4.32	0.0087
NM_016004	NM_016004	chromosome 20 open reading frame 9		0.2317	4.32	0.0087
NM_013236	NM_013236	like mouse brain protein E46		0.2318	4.31	0.0028
AK057677				0.2318	4.31	0.0046
NM_002906	NM_002906	radixin	cytoskeletal anchoring	0.2318	4.31	0.0055
NM_004805	NM_004805	DNA directed RNA polymerase II polypeptide D	transcription from Pol II promoter	0.2318	4.31	0.0080
NM_005390	NM_005390	pyruvate dehydrogenase (lipoamide) alpha 2	glycolysis; metabolism	0.2321	4.31	0.0021
AK001750	NM_018309	hypothetical protein FLJ11046		0.2322	4.31	0.0008
NM_002786	NM_002786; NM_148976	proteasome alpha 1 subunit isoform 2; proteasome alpha 1 subunit isoform 1	ubiquitin-dependent protein catabolism	0.2327	4.30	0.0043
NM_014487	NM_014487	zinc finger protein 330	electron transport; mitosis	0.2328	4.30	0.0068
NM_016078	NM_016078	CGI-148 protein		0.2333	4.29	0.0006
NM_003084	NM_003084	small nuclear RNA activating complex, polypeptide 3, 50kDa	snRNA transcription; transcription from Pol III promoter; regulation of transcription, DNA-dependent; transcription from Pol II promoter	0.2334	4.28	0.0037
NM_006357	NM_006357; NM_182678	ubiquitin-conjugating enzyme E2E 3	ubiquitin cycle	0.2340	4.27	0.0013
AJ243706	NM_006618	putative DNA/chromatin binding motif	oncogenesis	0.2340	4.27	0.0037
NM_006827	NM_006827	transmembrane trafficking protein	ER to Golgi transport; intracellular protein transport	0.2342	4.27	0.0004
BC008739	NM_020141	protein x 013		0.2342	4.27	0.0064
AK026980				0.2343	4.27	0.0135
AB051551		KIAA1764 protein		0.2343	4.27	0.0146
NM_024624	NM_024624	SMC6 protein		0.2344	4.27	0.0109
NM_003720	NM_003720	Down syndrome critical region protein 2		0.2347	4.26	0.0053
NM_012381	NM_012381; NM_181837	origin recognition complex, subunit 3 isoform 2; origin recognition complex, subunit 3 isoform 1	DNA replication	0.2351	4.25	0.0094
NM_013330	NM_013330; NM_197972	nucleoside-diphosphate kinase 7 isoform a; nucleoside-diphosphate kinase 7 isoform b	GTP biosynthesis; UTP biosynthesis; CTP biosynthesis	0.2353	4.25	0.0016
NM_031924	NM_031924	radial spokehead-like 2		0.2354	4.25	0.0095
NM_018838	NM_018838	13kDa differentiation-associated protein		0.2354	4.25	0.0088
NM_005552	NM_005552; NM_182923	kinesin 2 60/70kDa		0.2356	4.25	0.0063

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_004891	NM_004891; NM_145330	mitochondrial ribosomal protein L33 isoform a; mitochondrial ribosomal protein L33 isoform b	protein biosynthesis	0.2356	4.24	0.0050
NM_024047	NM_024047; NM_198038; NM_198039	nudix -type motif 9 isoform a; nudix -type motif 9 isoform b		0.2359	4.24	0.0091
NM_006758	NM_006758	U2 small nuclear RNA auxillary factor 1	RNA splicing; regulation of transcription, DNA-dependent; nuclear mRNA splicing, via spliceosome	0.2360	4.24	0.0109
NM_032179	NM_017871; NM_032179	hypothetical protein FLJ20542		0.2360	4.24	0.0174
NM_018685	NM_018685	anillin, actin binding protein (scraps homolog, *Drosophila*)	regulation of exit from mitosis; septin assembly and septum formation	0.2361	4.24	0.0008
BC008804	NM_015697	hypothetical protein CL640		0.2361	4.24	0.0173
AK057323	NM_182755	hypothetical protein LOC220929		0.2362	4.23	0.0066
NM_002069	NM_002069	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1	G-protein coupled receptor protein signaling pathway; signal transduction	0.2363	4.23	0.0037
AF070578	NM_153811	N system amino acid transporter NAT-1	amino acid transport; regulation of transcription, DNA-dependent	0.2365	4.23	0.0056
AK057477	NM_145014	hypothetical protein FLJ32915		0.2365	4.23	0.0051
NM_022457	NM_022457	constitutive photomorphogenic protein		0.2370	4.22	0.0069
NM_005614	NM_005614	Ras homolog enriched in brain	small GTPase mediated signal transduction	0.2371	4.22	0.0086
NM_004370	NM_004370; NM_080645	alpha 1 type XII collagen long isoform precursor; alpha 1 type XII collagen short isoform precursor	skeletal development	0.2375	4.21	0.0073
NM_003094	NM_003094	small nuclear ribonucleoprotein polypeptide E	spliceosome assembly	0.2375	4.21	0.0156
AL135926				0.2377	4.21	0.0085
BC014608	NM_138796	hypothetical protein BC014608		0.2378	4.21	0.0166
NM_000216	NM_000216	Kallmann syndrome 1 protein	axon guidance; chemotaxis; cell adhesion; cell motility	0.2382	4.20	0.0022
NM_015684	NM_015684	mitochondrial ATP synthase regulatory component factor B		0.2383	4.20	0.0015
NM_004792	NM_004792	peptidyl-prolyl isomerase G (cyclophilin G)	RNA splicing; protein folding	0.2385	4.19	0.0106
NM_003318	NM_003318	TTK protein kinase	mitotic spindle assembly; mitotic spindle checkpoint; positive regulation of cell proliferation; protein amino acid phosphorylation; regulation of cell cycle	0.2386	4.19	0.0017
NM_032133	NM_032133	Myc-binding protein-associated protein		0.2386	4.19	0.0133
NM_006531	NM_006531; NM_175605	Tg737 protein isoform 2; Tg737 protein isoform 1	excretion	0.2387	4.19	0.0026
NM_016262	NM_016262	tubulin, epsilon 1	microtubule-based movement; centrosome cycle	0.2389	4.19	0.0012
NM_005085	NM_005085	nucleoporin 214kDa	nucleocytoplasmic transport; oncogenesis	0.2390	4.18	0.0069
AF159141	NM_015399	breast cancer metastasis suppressor 1		0.2391	4.18	0.0049
NM_014713	NM_014713	lysosomal-associated protein transmembrane 4 alpha	transport	0.2391	4.18	0.0240
NM_015442	NM_015442	CCR4-NOT transcription complex, subunit 10		0.2392	4.18	0.0015
AK057730				0.2393	4.18	0.0006

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AF090094				0.2393	4.18	0.0191
AB018271	NM_001723; NM_015548; NM_020388; NM_183380	bullous pemphigoid antigen 1 isoform 1e precursor; bullous pemphigoid antigen 1 isoform 1eA precursor; bullous pemphigoid antigen 1 isoform 1eB precursor	cytoskeleton organization and biogenesis	0.2394	4.18	0.0003
NM_024580	NM_024580	hypothetical protein FLJ13119		0.2394	4.18	0.0071
NM_005025	NM_005025	serine (or cysteine) proteinase inhibitor, clade I (neuroserpin), member 1	peripheral nervous system development; central nervous system development	0.2394	4.18	0.0051
NM_012433	NM_012433	splicing factor 3b, subunit 1, 155kDa	nuclear mRNA splicing, via spliceosome	0.2396	4.17	0.0014
BC009561	NM_138787	hypothetical protein BC009561		0.2398	4.17	0.0072
NM_005038	NM_005038	peptidylprolyl isomerase D (cyclophilin D)	protein folding	0.2399	4.17	0.0049
NM_001813	NM_001813	centromere protein E	mitotic metaphase plate congression; mitotic chromosome movement; DNA replication and chromosome cycle; cytokinesis	0.2400	4.17	0.0009
NM_005534	NM_005534	interferon gamma receptor 2 (interferon gamma transducer 1)	response to pathogenic bacteria; response to virus; cell surface receptor linked signal transduction	0.2404	4.16	0.0010
NM_024619	NM_024619	hypothetical protein FLJ12171		0.2405	4.16	0.0186
NM_006811	NM_006811; NM_198941	tumor differentially expressed protein 1		0.2405	4.16	0.0025
NM_003620	NM_003620	protein phosphatase 1D	response to radiation; negative regulation of cell proliferation; regulation of cell cycle; protein amino acid dephosphorylation	0.2405	4.16	0.0094
NM_007192	NM_007192	chromatin-specific transcription elongation factor large subunit	nucleosome disassembly; proteolysis and peptidolysis; transcription from Pol II promoter	0.2409	4.15	0.0005
NM_014711	NM_014711	CP110 protein	biological_process unknown	0.2411	4.15	0.0118
NM_006246	NM_006246	epsilon isoform of regulatory subunit B56, protein phosphatase 2A	signal transduction	0.2411	4.15	0.0010
NM_002155	NM_002155	heat shock 70kDa protein 6 (HSP70B')	heat shock response	0.2411	4.15	0.0203
NM_019095	NM_019095	chromosome 20 open reading frame 155	phospholipid biosynthesis	0.2413	4.14	0.0043
BF541376				0.2414	4.14	0.0102
NM_012087	NM_012087	general transcription factor IIIC, polypeptide 5, 63kDa	transcription from Pol III promoter	0.2419	4.13	0.0140
NM_003797	NM_003797; NM_152991	embryonic ectoderm development isoform a; embryonic ectoderm development isoform b	negative regulation of transcription	0.2419	4.13	0.0016
BC002913	NM_052857	similar to RIKEN cDNA 2410003C20 gene		0.2420	4.13	0.0159
NM_021029	NM_021029	ribosomal protein L36a		0.2427	4.12	0.0154
NM_017832	NM_017832	hypothetical protein FLJ20457		0.2428	4.12	0.0007
BC011593	NM_138408	chromosome 6 open reading frame 51		0.2430	4.12	0.0067
NM_005527	NM_005527	heat shock 70kDa protein 1-like		0.2432	4.11	0.0062
NM_001892	NM_001892	casein kinase 1, alpha 1	protein amino acid phosphorylation; cell surface receptor linked signal transduction	0.2432	4.11	0.0000
AK057224	NM_152467	kelch-like 10		0.2435	4.11	0.0056
NM_001693	NM_001693	ATPase, H+ transporting, lysosomal 56/58kD, V1 subunit B, isoform 2	ATP biosynthesis; energy coupled proton transport, against the electrochemical gradient	0.2438	4.10	0.0030

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_004553	NM_004553	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)	mitochondrial electron transport, NADH to ubiquinone	0.2440	4.10	0.0093
NM_002337	NM_002337	low density lipoprotein receptor-related protein associated protein 1	vesicle-mediated transport; protein folding; cell proliferation	0.2440	4.10	0.0120
AB014560	NM_012297	Ras-GTPase activating protein SH3 domain-binding protein 2	cytoplasmic sequestering of NF-kappaB; protein-nucleus import; transport; RAS protein signal transduction	0.2443	4.09	0.0006
NM_003903	NM_003903	CDC16 homolog	regulation of mitosis; cell cycle; cytokinesis	0.2443	4.09	0.0020
NM_004953	NM_004953; NM_182917; NM_198241; NM_198242; NM_198244	eukaryotic translation initiation factor 4 gamma, 1 isoform 4; eukaryotic translation initiation factor 4 gamma, 1 isoform 1; eukaryotic translation initiation factor 4 gamma, 1 isoform 3; eukaryotic translation initiation factor 4 gamma, 1 isoform 2	regulation of translation	0.2446	4.09	0.0110
NM_032679	NM_032679	hypothetical protein MGC4400	regulation of transcription, DNA-dependent	0.2447	4.09	0.0018
AF042162				0.2449	4.08	0.0151
AB058765		KIAA1862 protein		0.2449	4.08	0.0039
NM_012385	NM_012385	p8 protein (candidate of metastasis 1)	induction of apoptosis; cell growth	0.2450	4.08	0.0044
NM_021994	NM_021994	zinc finger protein (C2H2 type) 277		0.2450	4.08	0.0001
NM_014962	NM_014962; NM_181443	BTB/POZ domain containing protein 3 isoform a; BTB/POZ domain containing protein 3 isoform b		0.2452	4.08	0.0004
BC004923	NM_033107	hypothetical protein BC004923		0.2453	4.08	0.0061
NM_031956	NM_031956	NYD-SP14 protein		0.2459	4.07	0.0005
NM_031371	NM_016374; NM_031371	RBP1-like protein 1 isoform 1; RBP1-like protein 1 isoform 2		0.2459	4.07	0.0089
AB020664	NM_015470	gamma-SNAP-associated factor 1		0.2464	4.06	0.0073
NM_032906	NM_032906	hypothetical protein MGC14156		0.2465	4.06	0.0097
NM_024688	NM_024688	hypothetical protein FLJ13031		0.2465	4.06	0.0022
AB007953				0.2467	4.05	0.0077
AF103803	NM_017548	hypothetical protein H41	cell proliferation	0.2467	4.05	0.0290
AK055310	NM_181756	zinc finger protein 233		0.2468	4.05	0.0005
NM_007278	NM_007278	GABA(A) receptor-associated protein	protein targeting; synaptic transmission	0.2469	4.05	0.0125
NM_001212	NM_001212	complement component 1, q subcomponent binding protein precursor	immune response	0.2472	4.05	0.0046
NM_002149	NM_002149; NM_134421	hippocalcin-like 1	vesicle-mediated transport	0.2472	4.04	0.0075
NM_004270	NM_004270	cofactor required for Sp1 transcriptional activation, subunit 9, 33kDa	transcription initiation from Pol II promoter; regulation of transcription from Pol II promoter	0.2473	4.04	0.0018
NM_003500	NM_003500	acyl-Coenzyme A oxidase 2, branched chain	bile acid metabolism; fatty acid beta-oxidation; fatty acid metabolism	0.2476	4.04	0.0050
NM_006021	NM_006021	deleted in lymphocytic leukemia, 2		0.2476	4.04	0.0493
NM_004652	NM_004652; NM_021906	ubiquitin specific protease 9, X-linked isoform 1; ubiquitin specific protease 9, X-linked isoform 2	deubiquitination; female gamete generation	0.2477	4.04	0.0123
NM_057180	NM_016226; NM_057180	vacuolar protein sorting 29 isoform 1; vacuolar protein sorting 29 isoform 2		0.2479	4.03	0.0040

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_007318	NM_000021; NM_007318; NM_007319	presenilin 1 isoform I-467; presenilin 1 isoform I-463; presenilin 1 isoform I-374	chromosome segregation; anti-apoptosis; chromosome organization and biogenesis (sensu Eukarya); intracellular signaling cascade	0.2479	4.03	0.0026
NM_014874	NM_014874	mitofusin 2		0.2481	4.03	0.0079
BC000365	NM_005316	general transcription factor IIH, polypeptide 1, 62kDa	regulation of CDK activity; regulation of transcription, DNA-dependent; DNA repair; transcription from Pol II promoter	0.2481	4.03	0.0057
NM_014390	NM_014390	staphylococcal nuclease domain containing 1		0.2482	4.03	0.0100
NM_001316	NM_001316; NM_177436	CSE1 chromosome segregation 1-like protein isoform a; CSE1 chromosome segregation 1-like protein isoform b	protein-nucleus import, docking; intracellular protein transport; apoptosis; cell proliferation	0.2485	4.02	0.0098
NM_031898	NM_031898	tektin 3	microtubule cytoskeleton organization and biogenesis	0.2486	4.02	0.0087
AK025719				0.2487	4.02	0.0051
NM_013446	NM_013446	makorin, ring finger protein, 1	biological_process unknown	0.2487	4.02	0.0005
BC017503		CDR2 protein	biological_process unknown	0.2487	4.02	0.0012
BC008952				0.2487	4.02	0.0012
NM_021188	NM_021188	clones 23667 and 23775 zinc finger protein	regulation of transcription, DNA-dependent	0.2488	4.02	0.0055
NM_021825	NM_021825	hypothetical protein MDS025		0.2491	4.01	0.0244
NM_000454	NM_000454	superoxide dismutase 1, soluble	superoxide metabolism; response to oxidative stress; neurogenesis	0.2499	4.00	0.0155
NM_030663	NM_030663	mitochondrial capsule selenoprotein	fertilization (sensu Animalia); sperm motility	0.2499	4.00	0.0352
NM_003010	NM_003010	mitogen-activated protein kinase kinase 4	JNK cascade; protein amino acid phosphorylation; signal transduction	0.2501	4.00	0.0015
NM_002157	NM_002157	heat shock 10kDa protein 1 (chaperonin 10)	protein folding	0.2501	4.00	0.0163
NM_014497	NM_014497	NP220 nuclear protein	mRNA splicing	0.2501	4.00	0.0170
NM_001398	NM_001398	peroxisomal enoyl-coenzyme A hydratase-like protein	fatty acid beta-oxidation; fatty acid metabolism; energy pathways	0.2505	3.99	0.0174
NM_006311	NM_006311	nuclear receptor co-repressor 1	transcription from Pol II promoter	0.2506	3.99	0.0189
NM_003472	NM_003472	DEK oncogene (DNA binding)	protein targeting; cell growth and/or maintenance; signal transduction; regulation of transcription from Pol II promoter; viral genome replication	0.2506	3.99	0.0093
AL117407	NM_015093; NM_145342	mitogen-activated protein kinase kinase kinase 7 interacting protein 2 isoform 1; mitogen-activated protein kinase kinase kinase 7 interacting protein 2 isoform 2		0.2507	3.99	0.0172
NM_016147	NM_016147	protein phosphatase methylesterase-1	protein amino acid demethylation	0.2513	3.98	0.0080
NM_022829	NM_022829	solute carrier family 13 member 3		0.2514	3.98	0.0162
NM_002372	NM_002372	mannosidase, alpha, class 2A, member 1	glycoprotein biosynthesis; carbohydrate metabolism	0.2516	3.97	0.0119
AK027647	NM_018981	ER-resident protein ERdj5		0.2517	3.97	0.0008
AB002308		KIAA0310 protein		0.2521	3.97	0.0005
NM_013341	NM_013341	hypothetical protein PTD004		0.2521	3.97	0.0065
NM_018255	NM_018255	elongator protein 2		0.2521	3.97	0.0013

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_006067	NM_006067	neighbor of COX4		0.2522	3.96	0.0085
NM_001011	NM_001011	ribosomal protein S7	protein biosynthesis	0.2522	3.96	0.0174
NM_032489	NM_032489	proacrosin binding protein sp32 precursor		0.2525	3.96	0.0102
NM_001449	NM_001449	four and a half LIM domains 1		0.2529	3.95	0.0022
NM_018230	NM_018230	nucleoporin 133kDa	transport; mRNA-nucleus export	0.2533	3.95	0.0081
AB025254	NM_014290	tudor repeat associator with PCTAIRE 2		0.2536	3.94	0.0214
AK058119	NM_145028	hypothetical protein FLJ25390		0.2537	3.94	0.0001
NM_014648	NM_014648	zinc finger DAZ interacting protein 3		0.2538	3.94	0.0018
NM_054016	NM_006625; NM_054016	FUS interacting protein (serine-arginine rich) 1 isoform 1; FUS interacting protein (serine-arginine rich) 1 isoform 2	mRNA processing; spliceosome assembly	0.2539	3.94	0.0012
NM_058004	NM_002650; NM_058004	phosphatidylinositol 4-kinase, catalytic, alpha polypeptide isoform 1; phosphatidylinositol 4-kinase, catalytic, alpha polypeptide isoform 2	phosphatidylinositol biosynthesis; synaptic transmission; signal transduction	0.2539	3.94	0.0106
NM_002647	NM_002647	phosphoinositide-3-kinase, class 3	non-selective vesicle transport	0.2540	3.94	0.0048
NM_006024	NM_006024	Tax1 (human T-cell leukemia virus type I) binding protein 1	anti-apoptosis	0.2544	3.93	0.0069
NM_018492	NM_018492	T-LAK cell-originated protein kinase		0.2545	3.93	0.0038
D87446				0.2545	3.93	0.0027
NM_003489	NM_003489	receptor interacting protein 140	transcription; regulation of transcription, DNA-dependent	0.2545	3.93	0.0016
NM_024629	NM_024629	KSHV latent nuclear antigen interacting protein 1		0.2547	3.93	0.0096
BC015621	NM_144578	MAPK-interacting and spindle-stabilizing protein		0.2547	3.93	0.0040
AK057021	NM_032661	hypothetical protein MGC5139		0.2548	3.92	0.0106
NM_012253	NM_012253	transketolase-like 1	thiamin metabolism; glucose catabolism	0.2550	3.92	0.0098
NM_016516	NM_016516	vacuolar protein sorting 54		0.2553	3.92	0.0029
NM_001867	NM_001867	cytochrome c oxidase subunit VIIc precursor	electron transport; energy pathways	0.2554	3.92	0.0082
NM_005762	NM_005762	tripartite motif-containing 28 protein	transcription regulation from Pol II promoter	0.2554	3.92	0.0009
AL050308				0.2555	3.91	0.0072
BC014605	NM_144627	SSTK-interacting protein		0.2556	3.91	0.0122
NM_032376	NM_032376	hypothetical protein MGC4251		0.2556	3.91	0.0063
NM_020404	NM_020404	tumor endothelial marker 1 precursor	biological_process unknown	0.2558	3.91	0.0115
NM_016146	NM_016146	trafficking protein particle complex 4	vesicle-mediated transport; dendrite morphogenesis; neurotransmitter receptor biosynthesis	0.2562	3.90	0.0089
BC015178	NM_015476	DKFZP586M1523 protein		0.2562	3.90	0.0023
NM_014236	NM_014236	glyceronephosphate O-acyltransferase	fatty acid metabolism; organogenesis	0.2563	3.90	0.0125
BC001854	NM_005911	methionine adenosyltransferase II, alpha	one-carbon compound metabolism	0.2564	3.90	0.0031
AL137562		hypothetical protein		0.2564	3.90	0.0161
NM_021911	NM_000813; NM_021911	gamma-aminobutyric acid (GABA) A receptor, beta 2 isoform 2; gamma-aminobutyric acid (GABA) A receptor, beta 2 isoform 1	gamma-aminobutyric acid signaling pathway; ion transport; synaptic transmission	0.2565	3.90	0.0093
NM_004208	NM_004208; NM_145812; NM_145813	programmed cell death 8 isoform 1; programmed cell death 8 isoform 2; programmed cell death 8 isoform 3	DNA damage response, signal transduction resulting in induction of apoptosis; DNA fragmentation; electron transport	0.2565	3.90	0.0134
AK023175	NM_017975	hypothetical protein FLJ10036		0.2565	3.90	0.0001

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_016576	NM_016576	guanosine monophosphate reductase 2		0.2565	3.90	0.0048
BC010744	NM_080677	dynein light chain 2	microtubule-based process	0.2571	3.89	0.0192
NM_025267	NM_025267	hypothetical protein MGC2744	alanyl-tRNA aminoacylation	0.2572	3.89	0.0092
NM_006303	NM_006303	JTV1	protein biosynthesis	0.2574	3.88	0.0090
AL136693	NM_024843	cytochrome b reductase 1	electron transport	0.2574	3.88	0.0022
NM_024532	NM_024532	PF20		0.2576	3.88	0.0071
NM_002835	NM_002835	protein tyrosine phosphatase, non-receptor type 12	protein amino acid dephosphorylation	0.2577	3.88	0.0003
NM_005134	NM_005134	protein phosphatase 4, regulatory subunit 1		0.2577	3.88	0.0056
X74794	NM_005914; NM_182746	minichromosome maintenance protein 4	DNA replication	0.2577	3.88	0.0070
AK023358				0.2578	3.88	0.0001
NM_003352	NM_003352	ubiquitin-like 1 (sentrin)	DNA repair	0.2578	3.88	0.0006
NM_001873	NM_001873	carboxypeptidase E precursor	neuropeptide signaling pathway; proteolysis and peptidolysis; metabolism; protein modification	0.2579	3.88	0.0117
NM_004965	NM_004965	high-mobility group nucleosome binding domain 1		0.2584	3.87	0.0221
NM_018114	NM_018114	hypothetical protein FLJ10496		0.2586	3.87	0.0024
NM_002526	NM_002526	5' nucleotidase, ecto	DNA metabolism; nucleotide catabolism	0.2586	3.87	0.0058
AF112219	NM_001984	esterase D/formylglutathione hydrolase	biological_process unknown	0.2588	3.86	0.0012
NM_014047	NM_014047	HSPC023 protein		0.2589	3.86	0.0204
AK057571	NM_014673	KIAA0103 gene product		0.2589	3.86	0.0030
NM_012158	NM_012158	F-box and leucine-rich repeat protein 3A	protein ubiquitination	0.2590	3.86	0.0022
NM_007266	NM_007266	XPA binding protein 1	small GTPase mediated signal transduction	0.2591	3.86	0.0022
NM_014338	NM_014338	phosphatidylserine decarboxylase	phospholipid biosynthesis	0.2592	3.86	0.0102
AL136807	NM_014445	stress-associated endoplasmic reticulum protein 1	plasma membrane organization and biogenesis; response to stress; protein amino acid glycosylation	0.2593	3.86	0.0037
NM_030790	NM_030790	T-cell immunomodulatory protein		0.2594	3.85	0.0064
BC016796		Unknown (protein for IMAGE:4070070)		0.2596	3.85	0.0017
AK026771	NM_020336	KIAA1219 protein		0.2596	3.85	0.0031
Z84469				0.2599	3.85	0.0035
NM_007372	NM_007372	RNA helicase-related protein		0.2600	3.85	0.0018
NM_003591	NM_003591	cullin 2	induction of apoptosis by intracellular signals; G1/S transition of mitotic cell cycle; cell cycle arrest; negative regulation of cell proliferation	0.2603	3.84	0.0114
NM_000126	NM_000126	electron transfer flavoprotein, alpha polypeptide	electron transport	0.2605	3.84	0.0103
NM_018281	NM_018281	hypothetical protein FLJ10948		0.2605	3.84	0.0093
BC008883	NM_025161	hypothetical protein FLJ22175		0.2608	3.83	0.0069
NM_013237	NM_013237	px19-like protein	development; immune response	0.2612	3.83	0.0060
NM_015907	NM_015907	leucine aminopeptidase	proteolysis and peptidolysis	0.2612	3.83	0.0076
NM_018289	NM_018289	hypothetical protein FLJ10979		0.2612	3.83	0.0033
NM_022362	NM_022362	MMS19-like (MET18 homolog, S. cerevisiae)		0.2612	3.83	0.0167
AL137257				0.2613	3.83	0.0034

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AB022663	NM_004290; NM_183398; NM_183399; NM_183400; NM_183401	ring finger protein 14 isoform 1; ring finger protein 14 isoform 2	ubiquitin cycle; signal transduction; regulation of transcription from Pol II promoter; androgen receptor signaling pathway	0.2618	3.82	0.0008
NM_006888	NM_006888	calmodulin 1	G-protein coupled receptor protein signaling pathway	0.2619	3.82	0.0097
NM_014791	NM_014791	maternal embryonic leucine zipper kinase	protein amino acid phosphorylation	0.2621	3.81	0.0044
NM_006003	NM_006003	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	electron transport	0.2623	3.81	0.0090
NM_024528	NM_024528	hypothetical protein FLJ22626	regulation of transcription, DNA-dependent	0.2625	3.81	0.0106
NM_004889	NM_004889	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit f, isoform 2	proton transport; ATP biosynthesis	0.2627	3.81	0.0107
NM_005617	NM_005617	ribosomal protein S14	protein biosynthesis	0.2629	3.80	0.0321
NM_016623	NM_016623	hypothetical protein BM-009		0.2631	3.80	0.0084
NM_032547	NM_032547	short coiled-coil protein		0.2634	3.80	0.0009
NM_007285	NM_007285	GABA(A) receptor-associated protein-like 2	intracellular protein transport	0.2636	3.79	0.0027
NM_031421	NM_031421	hypothetical protein DKFZp434H0115		0.2637	3.79	0.0140
NM_016008	NM_015522; NM_016008	dynein 2 light intermediate chain isoform 2; dynein 2 light intermediate chain isoform 1		0.2640	3.79	0.0002
NM_006055	NM_006055	lanthionine synthetase C-like protein 1	G-protein coupled receptor protein signaling pathway	0.2640	3.79	0.0021
AL161952	NM_002065	glutamate-ammonia ligase (glutamine synthase)	regulation of neurotransmitter levels; glutamine biosynthesis; nitrogen fixation	0.2641	3.79	0.0140
NM_014321	NM_014321	origin recognition complex subunit 6	DNA replication	0.2644	3.78	0.0000
NM_025211	NM_025211	protein kinase anchoring protein GKAP42		0.2646	3.78	0.0195
NM_020685	NM_020685	HT021		0.2648	3.78	0.0010
AF395440		HEJ1		0.2648	3.78	0.0130
AF204231	NM_015003; NM_181076; NM_181077	golgin-67 isoform a; golgin-67 isoform b; golgin-67 isoform c		0.2649	3.77	0.0025
NM_021096	NM_021096	calcium channel, voltage-dependent, alpha 1I subunit	cation transport; calcium ion transport	0.2649	3.77	0.0150
NM_003587	NM_003587	DEAH (Asp-Glu-Ala-His) box polypeptide 16	RNA splicing; regulation of cell cycle; nuclear mRNA splicing, via spliceosome	0.2654	3.77	0.0074
NM_016001	NM_016001	CGI-48 protein	regulation of transcription, DNA-dependent	0.2657	3.76	0.0046
NM_020194	NM_020194	GL004 protein		0.2660	3.76	0.0005
NM_017899	NM_017899	hypothetical protein FLJ20607		0.2661	3.76	0.0070
NM_005415	NM_005415	solute carrier family 20 (phosphate transporter), member 1	phosphate metabolism; small molecule transport	0.2662	3.76	0.0024
AL139112				0.2662	3.76	0.0138
AK054863				0.2663	3.75	0.0033
NM_016081	NM_016081	palladin		0.2664	3.75	0.0057
NM_012470	NM_012470	transportin 3	nucleocytoplasmic transport	0.2664	3.75	0.0149
AJ303380	NM_030906	serine/threonine kinase 33		0.2666	3.75	0.0086
Y08772		poly(A) binding protein		0.2666	3.75	0.0099
AL163279				0.2666	3.75	0.0018
NM_022473	NM_022473	zinc finger protein 106 homolog		0.2669	3.75	0.0002

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
AK025620				0.2672	3.74	0.0045
AK027135				0.2672	3.74	0.0125
NM_003516	NM_003516	H2A histone family, member O	nucleosome assembly; chromosome organization and biogenesis (sensu Eukarya)	0.2673	3.74	0.0039
AF042163				0.2675	3.74	0.0015
NM_000123	NM_000123	XPG-complementing protein	hearing; transcription-coupled nucleotide-excision repair	0.2678	3.73	0.0063
NM_018077	NM_018077	hypothetical protein FLJ10377		0.2680	3.73	0.0072
NM_015941	NM_015941	ATPase, H+ transporting, lysosomal 50/57kDa, V1 subunit H	proton transport; ATP biosynthesis	0.2681	3.73	0.0007
AF346509	NM_006599; NM_138713; NM_138714; NM_173214; NM_173215	nuclear factor of activated T-cells 5 isoform c; nuclear factor of activated T-cells 5 isoform b; nuclear factor of activated T-cells 5 isoform a	regulation of transcription, DNA-dependent; excretion; signal transduction; transcription from Pol II promoter	0.2683	3.73	0.0037
NM_024426	NM_000378; NM_024424; NM_024425; NM_024426	Wilms tumor 1 isoform A; Wilms tumor 1 isoform B; Wilms tumor 1 isoform C; Wilms tumor 1 isoform D	regulation of transcription, DNA-dependent; negative regulation of cell cycle	0.2685	3.72	0.0069
NM_018430	NM_018430	translin-associated factor X interacting protein 1		0.2687	3.72	0.0009
NM_015416	NM_015416	cervical cancer 1 protooncogene protein p40		0.2689	3.72	0.0043
NM_000693	NM_000693	aldehyde dehydrogenase 1A3	alcohol metabolism; lipid metabolism	0.2690	3.72	0.0082
AL133096	NM_018602	DnaJ (Hsp40) homolog, subfamily A, member 4		0.2690	3.72	0.0129
BC007436	NM_138369	hypothetical protein BC007436		0.2691	3.72	0.0092
BC015138	NM_152729	5'-nucleotidase, cytosolic II-like 1 protein		0.2692	3.72	0.0019
AB053314	NM_139163	ALS2CR12 gene product		0.2692	3.72	0.0005
AK024268			regulation of transcription, DNA-dependent	0.2692	3.71	0.0194
NM_016019	NM_016019	LUC7-like 2		0.2693	3.71	0.0016
NM_000310	NM_000310	palmitoyl-protein thioesterase 1 (ceroid-lipofuscinosis, neuronal 1, infantile)	vision; neurogenesis; protein modification	0.2696	3.71	0.0238
NM_002940	NM_002940	ATP-binding cassette, sub-family E, member 1	transport; electron transport; viral capsid assembly	0.2697	3.71	0.0007
NM_052969	NM_052969	ribosomal protein L39-like protein	protein biosynthesis	0.2697	3.71	0.0167
BC007973	NM_016534	apoptosis-related protein PNAS-1		0.2700	3.70	0.0050
NM_001985	NM_001985	electron-transfer-flavoprotein, beta polypeptide	electron transport	0.2701	3.70	0.0094
AK027191				0.2701	3.70	0.0145
NM_005605	NM_005605	protein phosphatase 3 (formerly 2B), catalytic subunit, gamma isoform (calcineurin A gamma)		0.2703	3.70	0.0098
BC015121				0.2704	3.70	0.0016
NM_017632	NM_017632	collaborates/cooperates with ARF (alternate reading frame) protein		0.2707	3.69	0.0029
NM_000039	NM_000039	apolipoprotein A-I precursor	lipid transport; cholesterol metabolism; circulation	0.2707	3.69	0.0043
NM_000089	NM_000089	alpha 2 type I collagen	skeletal development	0.2707	3.69	0.0102
AK026809	NM_145284	similar to hypothetical protein MGC17347		0.2711	3.69	0.0101
AB033076		KIAA1250 protein		0.2713	3.69	0.0016
NM_006370	NM_006370	vesicle-associated soluble NSF attachment protein receptor (v-SN	membrane fusion; vesicle transport; cell proliferation; non-selective vesicle docking	0.2714	3.68	0.0084

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_004800	NM_004800	transmembrane 9 superfamily member 2	transport	0.2715	3.68	0.0014
BC004895	NM_138368	hypothetical protein BC004895		0.2715	3.68	0.0143
NM_016122	NM_016122	NY-REN-58 antigen		0.2718	3.68	0.0004
NM_015965	NM_015965	cell death-regulatory protein GRIM19	protein-nucleus import; induction of apoptosis by extracellular signals; negative regulation of protein biosynthesis; negative regulation of cell growth; negative regulation of transcription, DNA-dependent; apoptotic nuclear changes	0.2718	3.68	0.0064
NM_005776	NM_005776	cornichon-like	intracellular signaling cascade; immune response	0.2718	3.68	0.0043
NM_024724	NM_024724	hypothetical protein FLJ22332		0.2719	3.68	0.0080
NM_002853	NM_002853; NM_133282; NM_133377	RAD1 homolog isoform 1; RAD1 homolog isoform 2	DNA damage response, signal transduction resulting in cell cycle arrest; meiotic prophase I; response to DNA damage stimulus; DNA repair	0.2719	3.68	0.0006
AL109684				0.2721	3.68	0.0004
AB037844		KIAA1423 protein		0.2723	3.67	0.0132
NM_032251	NM_032251	hypothetical protein DKFZp434G0920		0.2725	3.67	0.0082
NM_002162	NM_002162	intercellular adhesion molecule 3 precursor	cell-cell adhesion	0.2725	3.67	0.0047
NM_002097	NM_002097	general transcription factor IIIA	rRNA transcription; transcription from Pol III promoter; regulation of transcription, DNA-dependent	0.2726	3.67	0.0067
NM_001414	NM_001414	eukaryotic translation initiation factor 2B, subunit 1 alpha, 26kDa	translational initiation	0.2727	3.67	0.0113
NM_033022	NM_001026; NM_033022	ribosomal protein S24 isoform c; ribosomal protein S24 isoform a	protein biosynthesis	0.2732	3.66	0.0031
AB051504	NM_030648	SET domain-containing protein 7	chromatin modification	0.2733	3.66	0.0002
NM_003340	NM_003340; NM_181886; NM_181887; NM_181888; NM_181889; NM_181890; NM_181891; NM_181892; NM_181893	ubiquitin-conjugating enzyme E2D 3 isoform 1; ubiquitin-conjugating enzyme E2D 3 isoform 2; ubiquitin-conjugating enzyme E2D 3 isoform 3	protein modification; ubiquitin-dependent protein degradation	0.2733	3.66	0.0241
BC017696		Similar to RIKEN cDNA 2410075D05 gene		0.2733	3.66	0.0069
NM_014342	NM_014342	mitochondrial carrier homolog 2	transport	0.2733	3.66	0.0008
NM_020474	NM_020474	polypeptide N-acetylgalactosaminyltransferase 1	O-linked glycosylation; heterophilic cell adhesion	0.2733	3.66	0.0003
NM_006211	NM_006211	proenkephalin	neuropeptide signaling pathway; signal transduction; cell-cell signaling	0.2735	3.66	0.0220
AF148537	NM_007008; NM_020532; NM_153828	reticulon 4	apoptosis; negative regulation of axon extension; negative regulation of anti-apoptosis	0.2736	3.66	0.0106
NM_002492	NM_002492	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa precursor	mitochondrial electron transport, NADH to ubiquinone	0.2736	3.66	0.0095
AF255666	NM_015436	zinc finger protein 363		0.2736	3.65	0.0024

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_023012	NM_023012; NM_198261; NM_198262; NM_198263	similar to splicing factor, arginine/serine-rich 4 isoform a; similar to splicing factor, arginine/serine-rich 4 isoform b; similar to splicing factor, arginine/serine-rich 4 isoform c; similar to splicing factor, arginine/serine-rich 4 isoform d		0.2737	3.65	0.0048
NM_018210	NM_018210	hypothetical protein FLJ10769		0.2737	3.65	0.0039
NM_013360	NM_013360	zinc finger protein 222	regulation of transcription, DNA-dependent	0.2738	3.65	0.0023
AK027091				0.2739	3.65	0.0016
BC014000	NM_138447	hypothetical protein BC014000	regulation of transcription, DNA-dependent	0.2741	3.65	0.0017
NM_021631	NM_021631	apoptosis inhibitor	apoptosis	0.2742	3.65	0.0065
U61166		SH3 domain-containing protein SH3P17		0.2742	3.65	0.0005
NM_016056	NM_016056	CGI-119 protein		0.2743	3.65	0.0158
NM_018842	NM_018842	insulin receptor tyrosine kinase substrate		0.2745	3.64	0.0086
NM_001695	NM_001695	ATPase, H+ transporting, lysosomal 42kD, V1 subunit C, isoform 1	proton transport	0.2745	3.64	0.0133
NM_015361	NM_015361	R3H domain (binds single-stranded nucleic acids) containing		0.2746	3.64	0.0093
NM_004667	NM_004667	hect domain and RLD 2	ubiquitin cycle; intracellular protein transport	0.2747	3.64	0.0055
BC014603	NM_152779	hypothetical protein MGC26856	G-protein coupled receptor protein signaling pathway; vision	0.2748	3.64	0.0117
NM_014258	NM_014258	synaptonemal complex protein 2	synaptonemal complex formation; meiosis; cytokinesis	0.2748	3.64	0.0016
NM_016258	NM_016258	high glucose-regulated protein 8	humoral defense mechanism	0.2748	3.64	0.0175
NM_019005	NM_019005	hypothetical protein FLJ20323		0.2749	3.64	0.0045
NM_015339	NM_015339; NM_181442	activity-dependent neuroprotector	regulation of transcription, DNA-dependent	0.2750	3.64	0.0035
AC006017				0.2750	3.64	0.0005
NM_000919	NM_000919; NM_138766; NM_138821; NM_138822	peptidylglycine alpha-amidating monooxygenase isoform a, preproprotein; peptidylglycine alpha-amidating monooxygenase isoform b, preproprotein; peptidylglycine alpha-amidating monooxygenase isoform c, preproprotein; peptidylglycine alpha-amidating monooxygenase isoform d, preproprotein	protein modification	0.2750	3.64	0.0001
NM_005886	NM_005886	katanin p80 subunit B 1	cell cycle; cell motility; microtubule depolymerization	0.2752	3.63	0.0053
NM_004272	NM_004272	Homer, neuronal immediate early gene, 1B	metabotropic glutamate receptor, phospholipase C activating pathway; regulation of synapse	0.2752	3.63	0.0048
NM_001553	NM_001553	insulin-like growth factor binding protein 7	negative regulation of cell proliferation; regulation of cell growth	0.2753	3.63	0.0164
D87684	NM_014607	UBX domain containing 2	regulation of transcription, DNA-dependent	0.2753	3.63	0.0159
NM_022550	NM_003401; NM_022406; NM_022550	X-ray repair cross complementing protein 4 isoform 1; X-ray repair cross complementing protein 4 isoform 2	double-strand break repair; DNA recombination	0.2753	3.63	0.0027
NM_007208	NM_007208	mitochondrial ribosomal protein L3	protein biosynthesis	0.2756	3.63	0.0047
NM_020216	NM_020216	arginyl aminopeptidase (aminopeptidase B)	proteolysis and peptidolysis	0.2756	3.63	0.0182

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_058246	NM_005494; NM_058246	DnaJ (Hsp40) homolog, subfamily B, member 6 isoform b; DnaJ (Hsp40) homolog, subfamily B, member 6 isoform a	biological_process unknown	0.2757	3.63	0.0134
NM_018300	NM_018300	zinc finger protein 83 (HPF1)		0.2759	3.62	0.0055
NM_016462	NM_016462	chromosome 6 open reading frame 53		0.2760	3.62	0.0003
NM_014255	NM_014255	transmembrane protein 4		0.2760	3.62	0.0152
NM_016222	NM_016222	DEAD-box protein abstrakt	RNA processing; development; apoptosis	0.2761	3.62	0.0057
NM_005783	NM_005783	ATP binding protein associated with cell differentiation	electron transport	0.2763	3.62	0.0008
AB037735	NM_030672	KIAA1314 protein		0.2764	3.62	0.0232
BC015397	NM_138811	chromosome 7 open reading frame 31		0.2764	3.62	0.0015
NM_006567	NM_006567	phenylalanine-tRNA synthetase	tRNA processing; transport; phenylalanyl-tRNA aminoacylation	0.2765	3.62	0.0065
NM_017864	NM_017864	hypothetical protein FLJ20530		0.2766	3.62	0.0005
AF061735	NM_006356	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	proton transport	0.2766	3.62	0.0155
NM_003651	NM_003651	cold shock domain protein A	response to cold; negative regulation of transcription from Pol II promoter	0.2767	3.61	0.0181
AL133023	NM_021185	hypothetical protein DKFZp434A1022		0.2767	3.61	0.0023
NM_006063	NM_006063	sarcomeric muscle protein	striated muscle contraction	0.2768	3.61	0.0003
NM_005051	NM_005051	glutaminyl-tRNA synthetase	glutaminyl-tRNA aminoacylation; glutamyl-tRNA aminoacylation	0.2769	3.61	0.0048
NM_033657	NM_004632; NM_033657	death-associated protein 3	induction of apoptosis by extracellular signals	0.2769	3.61	0.0031
NM_002858	NM_002858	ATP-binding cassette, sub-family D, member 3	peroxisome organization and biogenesis; peroxisomal long-chain fatty acid import	0.2770	3.61	0.0013
NM_018474	NM_018474	uncharacterized hypothalamus protein HT013		0.2771	3.61	0.0088
AL353681				0.2773	3.61	0.0017
NM_006421	NM_006421	brefeldin A-inhibited guanine nucleotide-exchange protein 1	exocytosis	0.2773	3.61	0.0127
NM_021974	NM_021974	DNA directed RNA polymerase II polypeptide F	transcription from Pol II promoter	0.2776	3.60	0.0109
NM_006268	NM_006268	D4, zinc and double PHD fingers family 2	induction of apoptosis by extracellular signals; regulation of transcription, DNA-dependent	0.2777	3.60	0.0027
AL137660	NM_018051	hypothetical protein FLJ10300		0.2777	3.60	0.0077
NM_020154	NM_020154	chromosome 15 hypothetical ATG/GTP binding protein	biological_process unknown	0.2777	3.60	0.0148
NM_006645	NM_006645	START domain containing 10		0.2778	3.60	0.0094
AB020684		KIAA0877 protein	immune response	0.2780	3.60	0.0105
NM_000397	NM_000397	cytochrome b-245, beta polypeptide (chronic granulomatous disease)	electron transport; energy pathways; inflammatory response; antimicrobial humoral response (sensu Vertebrata)	0.2784	3.59	0.0077
NM_018364	NM_018364	hypothetical protein FLJ11220		0.2785	3.59	0.0004
NM_016207	NM_016207	cleavage and polyadenylation specific factor 3, 73kDa	mRNA processing	0.2786	3.59	0.0007

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_000587	NM_000587	complement component 7 precursor	response to pathogenic bacteria; complement activation, classical pathway; immune response; complement activation, alternative pathway; cytolysis	0.2791	3.58	0.0015
AB033114	NM_020749	KIAA1288 protein		0.2791	3.58	0.0014
NM_033328	NM_033328	capping protein alpha 3	actin cytoskeleton organization and biogenesis	0.2792	3.58	0.0059
NM_015510	NM_015510	DKFZP566O084 protein		0.2792	3.58	0.0040
X59405	NM_002389; NM_153826; NM_172350; NM_172351; NM_172352; NM_172353; NM_172354; NM_172355; NM_172356; NM_172357; NM_172358; NM_172359; NM_172360; NM_172361	membrane cofactor protein isoform 1 precursor; membrane cofactor protein isoform 4 precursor; membrane cofactor protein isoform 14 precursor; membrane cofactor protein isoform 3 precursor; membrane cofactor protein isoform 5 precursor; membrane cofactor protein isoform 6 precursor; membrane cofactor protein isoform 7 precursor; membrane cofactor protein isoform 9 precursor; membrane cofactor protein isoform 10 precursor; membrane cofactor protein isoform 11 precursor; membrane cofactor protein isoform 13 precursor; membrane cofactor protein isoform 2 precursor; membrane cofactor protein isoform 8 precursor; membrane cofactor protein isoform 12 precursor	virulence; complement activation; invasive growth	0.2793	3.58	0.0048
AK001731	NM_152261	hypothetical protein MGC17943		0.2797	3.58	0.0030
NM_005548	NM_005548	lysyl-tRNA synthetase	protein biosynthesis	0.2800	3.57	0.0063
NM_014737	NM_014737; NM_170773; NM_170774	Ras association domain family 2 isoform 1; Ras association domain family 2 isoform 2		0.2802	3.57	0.0025
BC000143	NM_022086; NM_133171; NM_182764	engulfment and cell motility 2	phagocytosis; apoptosis	0.2802	3.57	0.0005
NM_004939	NM_004939	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	ribosome biogenesis; cell growth and/or maintenance; spliceosome assembly; regulation of translational initiation; spermatogenesis; development	0.2804	3.57	0.0139
AL080113	NM_006386; NM_030881	DEAD box polypeptide 17 isoform p82; DEAD box polypeptide 17 isoform 2	RNA processing	0.2805	3.57	0.0041
NM_019038	NM_019038	tudor domain containing 4		0.2806	3.56	0.0117
NM_004362	NM_004362	calmegin	fertilization (sensu Animalia)	0.2807	3.56	0.0033
NM_017761	NM_017761	proline-rich nuclear receptor coactivator 2		0.2809	3.56	0.0004
NM_024692	NM_024692	hypothetical protein FLJ21069		0.2810	3.56	0.0022
AK023762	NM_003128; NM_178313	spectrin, beta, non-erythrocytic 1 isoform 1; spectrin, beta, non-erythrocytic 1 isoform 2		0.2810	3.56	0.0013
NM_005701	NM_005701	RNA, U transporter 1	nucleocytoplasmic transport	0.2810	3.56	0.0039
BC004538	NM_152271	hypothetical protein FLJ23749	ATP-dependent proteolysis	0.2811	3.56	0.0104
NM_031953	NM_031953	sorting nexin 25		0.2811	3.56	0.0042
AL050022	NM_015631	DKFZP564D116 protein		0.2811	3.56	0.0044
AL442092	NM_018334	leucine rich repeat neuronal 3		0.2812	3.56	0.0099
AL117595				0.2812	3.56	0.0049

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_002946	NM_002946	replication protein A2, 32kDa	DNA dependent DNA replication	0.2813	3.56	0.0054
NM_001010	NM_001010	ribosomal protein S6	protein biosynthesis	0.2813	3.55	0.0151
NM_006708	NM_006708	glyoxalase I	carbohydrate metabolism	0.2814	3.55	0.0062
NM_022157	NM_022157	Ras-related GTP binding C		0.2816	3.55	0.0196
AF049615		huntingtin interacting protein HYPM		0.2819	3.55	0.0128
NM_021980	NM_021980	optineurin	pathogenesis; cell death; signal transduction	0.2820	3.55	0.0024
NM_002078	NM_002078	golgi autoantigen, golgin subfamily a, 4	vesicle-mediated transport	0.2820	3.55	0.0064
AK056615	NM_152484	FLJ32053 protein		0.2820	3.55	0.0010
AF288392	NM_017673	chromosome 1 open reading frame 26		0.2821	3.55	0.0142
NM_032361	NM_032361	THO complex 3	transport; mRNA-nucleus export; nuclear mRNA splicing, via spliceosome	0.2821	3.54	0.0089
AB007935	NM_001542	immunoglobulin superfamily, member 3		0.2829	3.54	0.0252
NM_033028	NM_033028	Bardet-Biedl syndrome 4	vision	0.2831	3.53	0.0125
NM_014000	NM_003373; NM_014000	vinculin isoform VCL; VCL isoform meta-VCL	cell shape and cell size control	0.2831	3.53	0.0070
BC011751	NM_133491	polyamine N-acetyltransferase		0.2833	3.53	0.0077
NM_005494	NM_005494; NM_058246	DnaJ (Hsp40) homolog, subfamily B, member 6 isoform b; DnaJ (Hsp40) homolog, subfamily B, member 6 isoform a	biological_process unknown	0.2834	3.53	0.0136
NM_006839	NM_006839	inner membrane protein, mitochondrial (mitofilin)	biological_process unknown	0.2834	3.53	0.0153
NM_014477	NM_014477	chromosome 20 open reading frame 10	intracellular signaling cascade; negative regulation of cell growth	0.2834	3.53	0.0124
AJ420469				0.2840	3.52	0.0131
BC009905		Unknown (protein for MGC:2525)		0.2840	3.52	0.0049
NM_016171	NM_016171	prothymosin a14		0.2842	3.52	0.0109
NM_002074	NM_002074	guanine nucleotide-binding protein, beta-1 subunit	acetyl choline receptor signaling, muscarinic pathway; RAS protein signal transduction	0.2844	3.52	0.0127
NM_001107	NM_001107	erythrocyte acylphosphatase 1	phosphate metabolism	0.2846	3.51	0.0273
NM_003558	NM_003558	phosphatidylinositol-4-phosphate 5-kinase, type I, beta		0.2846	3.51	0.0001
Z29067	NM_002498; NM_152720	NIMA-related kinase 3	cell cycle; mitosis; protein amino acid phosphorylation; cytokinesis	0.2847	3.51	0.0111
NM_020401	NM_020401	nuclear pore complex protein	transport; mRNA-nucleus export	0.2848	3.51	0.0216
NM_003886	NM_003886; NM_139289	A-kinase anchor protein 4 isoform 1; A-kinase anchor protein 4 isoform 2	fertilization (sensu Animalia); cell motility; signal transduction	0.2848	3.51	0.0376
NM_015966	NM_015966; NM_198398	serologically defined breast cancer antigen 84 isoform b; serologically defined breast cancer antigen 84 isoform a		0.2849	3.51	0.0172
NM_023937	NM_023937	mitochondrial ribosomal protein L34	protein biosynthesis	0.2849	3.51	0.0119
NM_031423	NM_031423; NM_145697	cell division cycle associated 1		0.2850	3.51	0.0001
NM_003011	NM_003011	SET translocation (myeloid leukemia-associated)	nucleosome assembly; DNA replication; oncogenesis	0.2851	3.51	0.0148
AK025794	NM_014006; NM_015092	PI-3-kinase-related kinase SMG-1 isoform 2; PI-3-kinase-related kinase SMG-1 isoform 1		0.2851	3.51	0.0412
NM_019048	NM_019048	HCV NS3-transactivated protein 1	protein amino acid prenylation	0.2851	3.51	0.0098

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_020632	NM_020632; NM_130840; NM_130841	ATPase, H+ transporting, lysosomal V0 subunit a isoform 4		0.2852	3.51	0.0065
NM_000016	NM_000016	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	fatty acid beta-oxidation; electron transport; fatty acid metabolism; energy pathways	0.2853	3.51	0.0023
BC017179	NM_152455	hypothetical protein FLJ35867	regulation of transcription, DNA-dependent	0.2856	3.50	0.0142
NM_015896	NM_015896	zinc finger, MYND domain-containing 10		0.2860	3.50	0.0204
AL133102				0.2861	3.50	0.0021
NM_020359	NM_020359	phospholipid scramblase 2	phospholipid scrambling	0.2865	3.49	0.0088
NM_021145	NM_021145	cyclin D binding myb-like transcription factor 1		0.2866	3.49	0.0024
NM_004869	NM_004869	vacuolar protein sorting factor 4B	membrane fusion; regulation of transcription, DNA-dependent	0.2867	3.49	0.0036
NM_004247	NM_004247	U5 snRNP-specific protein, 116 kD	RNA splicing; translational elongation; nuclear mRNA splicing, via spliceosome	0.2868	3.49	0.0137
NM_024687	NM_024687	hypothetical protein FLJ23049		0.2868	3.49	0.0020
NM_032027	NM_032027	beta-amyloid binding protein precursor		0.2868	3.49	0.0006
NM_016329	NM_016329	Scm-like with four mbt domains 1		0.2874	3.48	0.0073
NM_031291	NM_031291	hypothetical protein DKFZp434N1235	transport	0.2874	3.48	0.0137
NM_033280	NM_033280	similar to signal peptidase complex (18kD)	proteolysis and peptidolysis; signal peptide processing	0.2874	3.48	0.0122
NM_017569	NM_017569	transcription factor (p38 interacting protein)		0.2876	3.48	0.0197
NM_006602	NM_006602	transcription factor-like 5 protein	spermatogenesis; cell proliferation; developmental processes; transcription from Pol II promoter	0.2877	3.48	0.0013
NM_007194	NM_007194; NM_145862	protein kinase CHK2 isoform a; protein kinase CHK2 isoform b		0.2878	3.47	0.0064
NM_032944	NM_031414; NM_032944	serine/threonine kinase 31 isoform a; serine/threonine kinase 31 isoform b	protein amino acid phosphorylation	0.2879	3.47	0.0175
NM_016020	NM_016020	transcription factor B1, mitochondrial	rRNA modification	0.2881	3.47	0.0197
AK056682				0.2881	3.47	0.0124
NM_032573	NM_032573	testis-specific protein TSP-NY		0.2882	3.47	0.0180
NM_005826	NM_005826	heterogeneous nuclear ribonucleoprotein R	mRNA processing; regulation of transcription, DNA-dependent	0.2886	3.47	0.0025
NM_018453	NM_018453	chromosome 14 open reading frame 11		0.2886	3.47	0.0021
NM_002087	NM_002087	granulin	positive regulation of cell proliferation; signal transduction; cell-cell signaling	0.2888	3.46	0.0170
AB058697		KIAA1794 protein		0.2890	3.46	0.0182
AK000745	NM_178812	LYRIC/3D3		0.2890	3.46	0.0025
NM_025075	NM_025075	hypothetical protein FLJ23445		0.2891	3.46	0.0033
AF301222				0.2893	3.46	0.0091
NM_016176	NM_016176; NM_016547	calcium binding protein Cab45 precursor		0.2894	3.46	0.0409
NM_030934	NM_030934	N2,N2-dimethylguanosine tRNA methyltransferase-like		0.2896	3.45	0.0236
NM_017746	NM_017746	hypothetical protein FLJ20287		0.2897	3.45	0.0002

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_007080	NM_007080	Sm protein F	RNA splicing; nuclear mRNA splicing, via spliceosome	0.2899	3.45	0.0122
NM_003582	NM_003582	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 3	protein amino acid phosphorylation	0.2900	3.45	0.0198
BC012356	NM_001126	adenylosuccinate synthase	AMP biosynthesis; purine nucleotide biosynthesis	0.2901	3.45	0.0033
AK054969	NM_174916	ubiquitin protein ligase E3 component n-recognin 1	ubiquitin cycle; protein biosynthesis	0.2903	3.44	0.0007
NM_001777	NM_001777; NM_198793	CD47 antigen isoform 1 precursor; CD47 antigen isoform 2 precursor	integrin-mediated signaling pathway; cell-matrix adhesion	0.2903	3.44	0.0001
NM_000314	NM_000314	phosphatase and tensin homolog	regulation of CDK activity; development; cell proliferation; protein amino acid dephosphorylation; negative regulation of cell cycle	0.2904	3.44	0.0029
NM_004426	NM_004426	polyhomeotic 1-like	'de novo' pyrimidine base biosynthesis; development	0.2906	3.44	0.0046
NM_024520	NM_024520	hypothetical protein FLJ22555		0.2907	3.44	0.0093
NM_018975	NM_018975	TRF2-interacting telomeric RAP1 protein	telomerase-dependent telomere maintenance	0.2908	3.44	0.0022
NM_000979	NM_000979	ribosomal protein L18	protein biosynthesis	0.2910	3.44	0.0069
NM_030924	NM_030924	bubblegum related protein	metabolism	0.2910	3.44	0.0169
NM_016505	NM_016505	putative S1 RNA binding domain protein		0.2912	3.43	0.0154
NM_000373	NM_000373	uridine monophosphate synthase	'de novo' pyrimidine base biosynthesis; UMP biosynthesis; nucleoside metabolism; pyrimidine nucleotide biosynthesis	0.2912	3.43	0.0100
NM_030877	NM_030877	beta catenin-like 1	regulation of transcription, DNA-dependent	0.2912	3.43	0.0006
NM_004681	NM_004681	eukaryotic translation initiation factor 1A, Y chromosome	translational initiation	0.2913	3.43	0.0033
NM_014394	NM_014394	growth hormone inducible transmembrane protein		0.2916	3.43	0.0193
AL354613	NM_018087	hypothetical protein FLJ10407		0.2916	3.43	0.0016
NM_002095	NM_002095	general transcription factor IIE, polypeptide 2, beta 34kDa	transcription initiation from Pol II promoter; regulation of transcription, DNA-dependent	0.2919	3.43	0.0114
NM_005901	NM_005901	MAD, mothers against decapentaplegic homolog 2	regulation of transcription, DNA-dependent; signal transduction	0.2920	3.42	0.0029
AB037795		KIAA1374 protein		0.2922	3.42	0.0019
AK055509				0.2924	3.42	0.0020
AL050005	NM_005801	putative translation initiation factor	cell growth and/or maintenance; regulation of translational initiation	0.2926	3.42	0.0188
NM_000999	NM_000999	ribosomal protein L38	protein biosynthesis	0.2929	3.41	0.0273
NM_004990	NM_004990	methionine-tRNA synthetase	protein biosynthesis	0.2930	3.41	0.0022
NM_017589	NM_017589	B-cell translocation gene 4		0.2931	3.41	0.0120
NM_001790	NM_001790; NM_022809	cell division cycle 25C protein isoform a; cell division cycle 25C protein isoform b	regulation of mitosis; regulation of CDK activity; start control point of mitotic cell cycle; protein amino acid dephosphorylation; cytokinesis	0.2931	3.41	0.0005
NM_020987	NM_001149; NM_020987	ankyrin 3 isoform 2; ankyrin 3 isoform 1	cytoskeletal anchoring; nonselective vesicle transport	0.2931	3.41	0.0043

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_017577	NM_017577	hypothetical protein DKFZp434C0328		0.2933	3.41	0.0005
NM_001464	NM_001464	a disintegrin and metalloproteinase domain 2 proprotein	fusion of sperm to egg plasma membrane; proteolysis and peptidolysis; cell adhesion	0.2934	3.41	0.0096
NM_003234	NM_003234	transferrin receptor (p90, CD71)	iron transport; iron homeostasis	0.2936	3.41	0.0061
AL137077				0.2937	3.41	0.0100
NM_018436	NM_018436	allantoicase isoform a		0.2937	3.40	0.0037
NM_002894	NM_002894	retinoblastoma binding protein 8	DNA repair; cell cycle checkpoint; transcription regulation from Pol II promoter	0.2939	3.40	0.0005
AB032981		KIAA1155 protein		0.2940	3.40	0.0143
NM_002584	NM_002584; NM_013945	paired box gene 7 isoform 1; paired box gene 7 isoform 2	morphogenesis; anti-apoptosis; cell growth and/or maintenance; regulation of transcription, DNA-dependent	0.2943	3.40	0.0002
BC000233	NM_005272	guanine nucleotide binding protein, alpha transducing activity polypeptide 2	phototransduction; G-protein coupled receptor protein signaling pathway; vision; signal transduction	0.2943	3.40	0.0113
NM_019896	NM_019896	polymerase (DNA-directed), epsilon 4 (p12 subunit)		0.2943	3.40	0.0036
BC009702	NM_145261	similar to RIKEN cDNA 1810055D05		0.2944	3.40	0.0009
NM_003157	NM_003157	NIMA (never in mitosis gene a)-related kinase 4	mitosis; protein amino acid phosphorylation	0.2944	3.40	0.0181
AB033113	NM_020748	KIAA1287 protein		0.2948	3.39	0.0012
NM_014549	NM_014549	POM121-like protein	regulation of transcription, DNA-dependent	0.2952	3.39	0.0089
AK054769	NM_006402	hepatitis B virus x-interacting protein	viral genome replication	0.2952	3.39	0.0138
NM_003115	NM_003115	UDP-N-acteylglucosamine pyrophosphorylase 1	UDP-N-acetylglucosamine biosynthesis; metabolism	0.2953	3.39	0.0098
NM_024040	NM_024040	hypothetical protein MGC2491		0.2955	3.38	0.0248
NM_018675	NM_018443; NM_018675	zinc finger protein 302; ZNF135-like protein	regulation of transcription, DNA-dependent	0.2957	3.38	0.0103
NM_032520	NM_032520	hypothetical protein CAB56184		0.2958	3.38	0.0079
AK057305	NM_145020	hypothetical protein FLJ32743		0.2959	3.38	0.0202
NM_003016	NM_003016	splicing factor, arginine/serine-rich 2	mRNA splicing; mRNA processing	0.2959	3.38	0.0156
NM_005433	NM_005433	viral oncogene yes-1 homolog 1	cell growth and/or maintenance; protein amino acid phosphorylation; intracellular signaling cascade	0.2960	3.38	0.0120
AB040882	NM_020839	WD repeat endosomal protein		0.2961	3.38	0.0131
NM_004788	NM_004788	ubiquitination factor E4A	ubiquitin-dependent protein catabolism	0.2962	3.38	0.0212
NM_020362	NM_020362	HT014		0.2964	3.37	0.0007
NM_014052	NM_014052	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide		0.2965	3.37	0.0180
NM_004619	NM_004619; NM_145759	TNF receptor-associated factor 5	apoptosis; signal transduction	0.2965	3.37	0.0014
NM_005003	NM_005003	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8kDa	fatty acid biosynthesis	0.2965	3.37	0.0040
NM_002093	NM_002093	glycogen synthase kinase 3 beta	glycogen metabolism; protein amino acid phosphorylation	0.2965	3.37	0.0090

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_000841	NM_000841	glutamate receptor, metabotropic 4	negative regulation of adenylate cyclase activity; G-protein coupled receptor protein signaling pathway; synaptic transmission	0.2966	3.37	0.0069
NM_022818	NM_022818	microtubule-associated proteins 1A/1B light chain 3		0.2967	3.37	0.0144
AF155654	NM_152301	hypothetical protein MGC9651		0.2967	3.37	0.0054
NM_006787		melanoma antigen, family D, 2		0.2968	3.37	0.0038
NM_022826	NM_022826	axotrophin		0.2969	3.37	0.0051
NM_016470	NM_016470	chromosome 20 open reading frame 111		0.2970	3.37	0.0279
AB002323	NM_001376	dynein, cytoplasmic, heavy polypeptide 1	mitotic spindle assembly; microtubule-based movement	0.2971	3.37	0.0196
X60155	NM_007130; NM_153380	zinc finger protein 41	regulation of transcription, DNA-dependent; biological_process unknown	0.2972	3.36	0.0041
NM_005066	NM_005066	splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)	RNA splicing; nuclear mRNA splicing, via spliceosome	0.2975	3.36	0.0417
AK057543	NM_130899	hypothetical protein MGC26988		0.2975	3.36	0.0272
NM_016488	NM_016488	periphilin 1		0.2976	3.36	0.0020
NM_006577	NM_006577; NM_033252	beta-1,3-N-acetylglucosaminyltransferase bGnT-1	protein amino acid glycosylation; biological_process unknown	0.2976	3.36	0.0355
NM_018256	NM_018256	WD repeat domain 12 protein		0.2976	3.36	0.0020
NM_007015	NM_007015	chondromodulin I precursor	proteoglycan metabolism; cell growth and/or maintenance; skeletal development	0.2980	3.36	0.0016
NM_003630	NM_003630	peroxisomal biogenesis factor 3	peroxisome organization and biogenesis	0.2981	3.35	0.0069
BC013151	NM_138451	hypothetical protein BC013151	regulation of transcription, DNA-dependent	0.2985	3.35	0.0034
NM_014177	NM_014177	HSPC154 protein		0.2986	3.35	0.0057
AK055091				0.2986	3.35	0.0055
NM_014931	NM_014931	KIAA1115 protein		0.2988	3.35	0.0021
NM_032320	NM_032320	K+ channel tetramerization protein		0.2989	3.35	0.0075
NM_019058	NM_019058	RTP801		0.2989	3.35	0.0260
D86981	NM_006380	amyloid beta precursor protein-binding protein 2	intracellular protein transport	0.2990	3.34	0.0073
AF085233	NM_013257; NM_170709	serum/glucocorticoid regulated kinase-like isoform 1; serum/glucocorticoid regulated kinase-like isoform 2	protein amino acid phosphorylation; intracellular signaling cascade	0.2990	3.34	0.0049
AB029041		KIAA1118 protein		0.2991	3.34	0.0189
AK023151	NM_024953	hypothetical protein FLJ13089		0.2991	3.34	0.0029
NM_006500	NM_006500	melanoma cell adhesion molecule	morphogenesis; cell adhesion	0.2993	3.34	0.0010
NM_014050	NM_014050; NM_172177; NM_172178	mitochondrial ribosomal protein L42 isoform a; mitochondrial ribosomal protein L42 isoform b	protein biosynthesis	0.2995	3.34	0.0159
AB058747	NM_016628; NM_100264; NM_100486	WW domain-containing adapter with a coiled-coil region isoform 1; WW domain-containing adapter with a coiled-coil region isoform 2; WW domain-containing adapter with a coiled-coil region isoform 3		0.2995	3.34	0.0004
NM_016291	NM_016291	inositol hexaphosphate kinase 2		0.2995	3.34	0.0054
X89657				0.2996	3.34	0.0023

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_018446	NM_018446; NM_152932	glycosyltransferase AD-017		0.2996	3.34	0.0061
NM_006940	NM_006940; NM_152989; NM_178010	SRY (sex determining region Y)-box 5 isoform a; SRY (sex determining region Y)-box 5 isoform b; SRY (sex determining region Y)-box 5 isoform c	transcription from Pol II promoter	0.2998	3.34	0.0058
NM_017760	NM_017760	hypothetical protein FLJ20311		0.2998	3.34	0.0006
NM_032988	NM_012453; NM_032988	transducin (beta)-like 2 isoform 1; transducin (beta)-like 2 isoform 2	biological_process unknown	0.2998	3.34	0.0271
NM_004666	NM_004666	vanin 1 precursor	cell motility; nitrogen metabolism	0.2998	3.34	0.0061
AK021841				0.2998	3.34	0.0027
AK058080	NM_152525	hypothetical protein FLJ25351		0.2999	3.33	0.0097
NM_031217	NM_031217	kinesin family member 18A		0.2999	3.33	0.0075
NM_004127	NM_004127	G protein pathway suppressor 1	JNK cascade; inactivation of MAPK; cell cycle	0.2999	3.33	0.0119
NM_002815	NM_002815	proteasome 26S non-ATPase subunit 11		0.3001	3.33	0.0066
AK002085				0.3003	3.33	0.0004
NM_006903	NM_006903; NM_176866; NM_176867; NM_176869	inorganic pyrophosphatase 2 isoform 2; inorganic pyrophosphatase 2 isoform 3; inorganic pyrophosphatase 2 isoform 4; inorganic pyrophosphatase 2 isoform 1		0.3003	3.33	0.0017
NM_005750	NM_005750	chromosome 4 open reading frame 6	neurogenesis	0.3004	3.33	0.0210
NM_016410	NM_016410	hypothetical protein HSPC177		0.3004	3.33	0.0118
NM_016936	NM_016936	ubinuclein 1	regulation of transcription from Pol II promoter	0.3006	3.33	0.0050
AL096749	NM_198547	FLJ46354 protein		0.3006	3.33	0.0160
NM_002129	NM_002129	high-mobility group box 2	regulation of transcription, DNA-dependent	0.3007	3.33	0.0428
BC011587		Similar to RIKEN cDNA 1700018O18 gene		0.3010	3.32	0.0120
NM_005811	NM_005811	growth differentiation factor 11	mesoderm development; skeletal development; neurogenesis; growth	0.3010	3.32	0.0081
AK024951	NM_001733	complement component 1, r subcomponent	complement activation, classical pathway; proteolysis and peptidolysis; immune response	0.3011	3.32	0.0209
NM_014166	NM_014166	vitamin D receptor interacting protein	transcription initiation from Pol II promoter; regulation of transcription, DNA-dependent; androgen receptor signaling pathway	0.3012	3.32	0.0020
NM_007207	NM_007207; NM_144728; NM_144729	dual specificity phosphatase 10 isoform a; dual specificity phosphatase 10 isoform b	JNK cascade; response to stress; protein amino acid dephosphorylation	0.3012	3.32	0.0380
AK021551				0.3012	3.32	0.0015
AK025557				0.3012	3.32	0.0065
NM_000386	NM_000386	bleomycin hydrolase	proteolysis and peptidolysis	0.3012	3.32	0.0020
AF190642	NM_016341	pancreas-enriched phospholipase C		0.3013	3.32	0.0002
NM_013285	NM_013285	nucleolar GTPase	biological_process unknown	0.3013	3.32	0.0018
AL133101				0.3015	3.32	0.0241
NM_006918	NM_006918	sterol-C5-desaturase (ERG3 delta-5-desaturase homolog, fungal)-like	lipid metabolism; sterol biosynthesis	0.3019	3.31	0.0176
AK057296	NM_144681	hypothetical protein FLJ32734		0.3021	3.31	0.0111

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_003404	NM_003404; NM_139323	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide		0.3022	3.31	0.0157
NM_005713	NM_005713; NM_031361	alpha 3 type IV collagen binding protein isoform 1; alpha 3 type IV collagen binding protein isoform 2	protein amino acid phosphorylation; immune response	0.3023	3.31	0.0125
NM_021999	NM_021999	integral membrane protein 2B	hearing; neurogenesis	0.3024	3.31	0.0233
AK055891	NM_024549	hypothetical protein FLJ21127		0.3027	3.30	0.0000
NM_020244	NM_020244	choline phosphotransferase 1	phospholipid biosynthesis; electron transport	0.3029	3.30	0.0019
NM_014937	NM_014937; NM_198330; NM_198331	inositol polyphosphate-5-phosphatase F isoform 1; inositol polyphosphate-5-phosphatase F isoform 2; inositol polyphosphate-5-phosphatase F isoform 3		0.3030	3.30	0.0016
NM_030594	NM_030594	cytoplasmic polyadenylation element binding protein 1		0.3030	3.30	0.0018
NM_020187	NM_020187	DC12 protein		0.3031	3.30	0.0027
NM_006325	NM_006325	ras-related nuclear protein	mitotic spindle assembly; protein-nucleus export; small GTPase mediated signal transduction; RNA-nucleus export; intracellular protein transport; regulation of cell cycle; DNA metabolism	0.3032	3.30	0.0352
BC014941	NM_001546	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein	regulation of transcription from Pol II promoter	0.3033	3.30	0.0034
NM_015959	NM_015959	thioredoxin-related transmembrane protein 2	electron transport	0.3034	3.30	0.0042
NM_017867	NM_017867	hypothetical protein FLJ20534		0.3034	3.30	0.0076
AK025774	NM_018206	vacuolar protein sorting 35	intracellular protein transport; retrograde (endosome to Golgi) transport	0.3037	3.29	0.0319
NM_014364	NM_014364	glyceraldehyde-3-phosphate dehydrogenase, testis-specific	glycolysis	0.3038	3.29	0.0074
NM_004457	NM_004457	long-chain fatty-acid-Coenzyme A ligase 3		0.3042	3.29	0.0117
NM_004330	NM_004330	BCL2/adenovirus E1B 19kD interacting protein 2	anti-apoptosis	0.3045	3.28	0.0112
NM_015370	NM_015370	hypothetical protein HS747E2A		0.3046	3.28	0.0347
AK025583				0.3046	3.28	0.0066
AB033029	NM_020718	ubiquitin-specific protease KIAA1203	ubiquitin-dependent protein catabolism	0.3047	3.28	0.0022
BC013920				0.3052	3.28	0.0138
NM_014500	NM_014500	HIV TAT specific factor 1	regulation of transcription from Pol II promoter; viral genome replication	0.3052	3.28	0.0151
AK026646	NM_033161	surfeit 4	biological_process unknown	0.3053	3.28	0.0003
NM_014239	NM_014239	eukaryotic translation initiation factor 2B, subunit 2 beta, 39kDa	translational initiation	0.3054	3.27	0.0049
NM_014924	NM_014924	KIAA0831		0.3055	3.27	0.0154
BC018085	NM_145267	chromosome 6 open reading frame 57		0.3055	3.27	0.0024
AK022842				0.3058	3.27	0.0008
NM_016200	NM_016200	U6 snRNA-associated Sm-like protein LSm8	nuclear mRNA splicing, via spliceosome	0.3059	3.27	0.0138
BC001563	NM_144596; NM_198309; NM_198310	tetratricopeptide repeat domain 8 isoform 3; tetratricopeptide repeat domain 8 isoform 1; tetratricopeptide repeat domain 8 isoform 2		0.3059	3.27	0.0041

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_006867	NM_006867	RNA-binding protein with multiple splicing	RNA processing	0.3059	3.27	0.0177
NM_006330	NM_006330	lysophospholipase I		0.3060	3.27	0.0030
NM_015285	NM_015285; NM_052834	WD repeat domain 7 protein isoform 1; WD repeat domain 7 protein isoform 2		0.3060	3.27	0.0015
NM_024899	NM_024899	hypothetical protein FLJ12542		0.3061	3.27	0.0081
NM_006437	NM_006437	poly(ADP-ribosyl)transferase-like 1	protein amino acid ADP-ribosylation; transport; necrosis; DNA repair; inflammatory response; response to drug	0.3061	3.27	0.0099
NM_032126	NM_032126	hypothetical protein DKFZp564J047		0.3063	3.26	0.0153
NM_001821	NM_001821	choroideremia-like Rab escort protein 2	vision; intracellular protein transport	0.3072	3.26	0.0021
NM_002402	NM_002402; NM_177524; NM_177525	mesoderm specific transcript isoform a; mesoderm specific transcript isoform b	mesoderm development	0.3073	3.25	0.0031
AB051537	NM_033512	KIAA1750 protein	nucleosome assembly; regulation of transcription, DNA-dependent	0.3074	3.25	0.0042
M12679	NM_002117	major histocompatibility complex, class I, C precursor	immune response	0.3075	3.25	0.0102
NM_016052	NM_016052	CGI-115 protein		0.3076	3.25	0.0330
AY007149				0.3076	3.25	0.0079
NM_002692	NM_002692	polymerase (DNA directed), epsilon 2 (p59 subunit)	DNA replication; DNA repair	0.3079	3.25	0.0159
NM_016648	NM_016648	HDCMA18P protein		0.3081	3.25	0.0061
NM_003457	NM_003457	zinc finger protein 207	regulation of transcription, DNA-dependent	0.3083	3.24	0.0125
AK025656				0.3084	3.24	0.0087
BC009230	NM_130467	PAGE-5 protein		0.3086	3.24	0.0238
NM_030974	NM_030974	hypothetical protein DKFZp434N1923		0.3087	3.24	0.0133
NM_032837	NM_032837	hypothetical protein FLJ14775	regulation of transcription, DNA-dependent	0.3087	3.24	0.0146
AF041429	NM_145169	chromosome 6 open reading frame 83		0.3087	3.24	0.0014
AL136945	NM_005455	zinc finger protein 265	RNA splicing	0.3088	3.24	0.0259
NM_005484	NM_005484	poly (ADP-ribosyl) transferase-like 2	protein amino acid ADP-ribosylation	0.3089	3.24	0.0019
NM_018442	NM_018442	PC326 protein		0.3090	3.24	0.0238
NM_004083	NM_004083	DNA-damage-inducible transcript 3	cell cycle arrest; response to DNA damage stimulus; cell growth and/or maintenance; regulation of transcription, DNA-dependent	0.3091	3.24	0.0158
D29954			regulation of transcription, DNA-dependent	0.3092	3.23	0.0258
NM_018840	NM_018840	putative Rab5-interacting protein		0.3093	3.23	0.0038
AK057629				0.3093	3.23	0.0072
NM_001064	NM_001064	transketolase		0.3094	3.23	0.0072
NM_003611	NM_003611	oral-facial-digital syndrome 1	defense response; biological_process unknown	0.3095	3.23	0.0170
NM_003757	NM_003757	eukaryotic translation initiation factor 3, subunit 2 beta, 36kDa	regulation of translational initiation; protein biosynthesis	0.3095	3.23	0.0093
NM_019006	NM_019006	protein associated with PRK1		0.3095	3.23	0.0150

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_031370	NM_002138; NM_031369; NM_031370	heterogeneous nuclear ribonucleoprotein D isoform c; heterogeneous nuclear ribonucleoprotein D isoform b; heterogeneous nuclear ribonucleoprotein D isoform a	RNA processing; RNA catabolism	0.3096	3.23	0.0095
AL137295	NM_004641	myeloid/lymphoid or mixed-lineage leukemia (trithorax (Drosophil	cell growth and/or maintenance; regulation of transcription, DNA-dependent	0.3097	3.23	0.0252
NM_002134	NM_002134	heme oxygenase (decyclizing) 2	heme oxidation	0.3097	3.23	0.0169
NM_003878	NM_003878	gamma-glutamyl hydrolase precursor		0.3100	3.23	0.0013
Z82202				0.3101	3.23	0.0190
AC004955				0.3103	3.22	0.0072
NM_032231	NM_032231	hypothetical protein FLJ22875		0.3103	3.22	0.0003
NM_004504	NM_004504	HIV-1 Rev binding protein	transport; mRNA-nucleus export	0.3105	3.22	0.0145
BC006436	NM_032497	hypothetical protein MGC13105	regulation of transcription, DNA-dependent	0.3106	3.22	0.0038
NM_016039	NM_016039	chromosome 14 open reading frame 166		0.3106	3.22	0.0174
NM_005087	NM_005087	fragile X mental retardation-related protein 1	regulation of transcription, DNA-dependent; apoptosis	0.3107	3.22	0.0036
NM_003145	NM_003145	signal sequence receptor, beta precursor	cotranslational membrane targeting	0.3107	3.22	0.0038
BC017590		Unknown (protein for MGC:26890)		0.3109	3.22	0.0308
NM_016626	NM_016626	hypothetical protein LOC51320		0.3110	3.22	0.0014
NM_007044	NM_007044	katanin p60 subunit A 1	mitosis; cell shape and cell size control	0.3112	3.21	0.0148
AF070641				0.3112	3.21	0.0030
NM_018159	NM_018159	nudix (nucleoside diphosphate linked moiety X)-type motif 11		0.3115	3.21	0.0232
NM_006571	NM_006571	dynactin 6	mitochondrion organization and biogenesis; lipid biosynthesis	0.3115	3.21	0.0095
NM_001998	NM_001998	fibulin 2 precursor		0.3116	3.21	0.0190
NM_002306	NM_002306	galectin-3	heterophilic cell adhesion	0.3119	3.21	0.0058
NM_030809	NM_030809	TGF-beta induced apotosis protein 12		0.3121	3.20	0.0002
AL080132				0.3121	3.20	0.0046
NM_005918	NM_005918	mitochondrial malate dehydrogenase precursor	tricarboxylic acid cycle; biological_process unknown	0.3122	3.20	0.0044
BC017472	NM_152682	hypothetical protein MGC10198		0.3123	3.20	0.0036
AK057395	NM_152488	hypothetical protein FLJ32833		0.3123	3.20	0.0136
AF155103	NM_033121	ankyrin repeat domain 13		0.3123	3.20	0.0247
BC015365	NM_145274	hypothetical protein MGC21518		0.3126	3.20	0.0006
AK021668				0.3126	3.20	0.0091
NM_004356	NM_004356	CD81 antigen	defense response; cell proliferation; protein complex assembly	0.3129	3.20	0.0091
NM_018571	NM_018571	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 2	protein amino acid phosphorylation	0.3130	3.20	0.0003
NM_014803		KIAA0335 gene product		0.3132	3.19	0.0027
NM_014642	NM_014642	KIAA0036 gene product		0.3132	3.19	0.0250
AL359591	NM_018710	hypothetical protein DKFZp762O076		0.3133	3.19	0.0009
NM_003800	NM_003800	RNA guanylyltransferase and 5'-phosphatase	mRNA capping; protein amino acid dephosphorylation	0.3133	3.19	0.0010
NM_000287	NM_000287	peroxisomal biogenesis factor 6	protein-peroxisome targeting; peroxisome organization and biogenesis	0.3135	3.19	0.0092

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_031420	NM_031420	mitochondrial ribosomal protein L9	protein biosynthesis	0.3135	3.19	0.0057
NM_024586	NM_024586; NM_148904; NM_148905; NM_148906; NM_148907; NM_148908; NM_148909	oxysterol-binding protein-like protein 9 isoform e; oxysterol-binding protein-like protein 9 isoform a; oxysterol-binding protein-like protein 9 isoform b; oxysterol-binding protein-like protein 9 isoform c; oxysterol-binding protein-like protein 9 isoform d; oxysterol-binding protein-like protein 9 isoform f		0.3138	3.19	0.0087
AK057666				0.3140	3.18	0.0026
BC015370		Unknown (protein for IMAGE:3916023)		0.3141	3.18	0.0073
AF278605				0.3143	3.18	0.0098
NM_017906	NM_017906	PAK1 interacting protein 1		0.3144	3.18	0.0077
NM_002495	NM_002495	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	mitochondrial electron transport, NADH to ubiquinone	0.3146	3.18	0.0111
NM_001923	NM_001923	damage-specific DNA binding protein 1	nucleotide-excision repair	0.3150	3.17	0.0031
NM_015640	NM_015640	PAI-1 mRNA-binding protein	regulation of transcription, DNA-dependent	0.3150	3.17	0.0279
NM_004092	NM_004092	mitochondrial short-chain enoyl-coenzyme A hydratase 1 precursor	fatty acid beta-oxidation; fatty acid metabolism; energy pathways	0.3150	3.17	0.0075
BC014310	NM_080746	ribosomal protein L10-like protein	protein biosynthesis	0.3151	3.17	0.0019
NM_031966	NM_031966	cyclin B1	G2/M transition of mitotic cell cycle; mitosis; regulation of cell cycle; cytokinesis	0.3151	3.17	0.0457
BC011706	NM_144988	hypothetical protein MGC19780		0.3152	3.17	0.0103
NM_002621	NM_002621	properdin P factor, complement	immune response; complement activation, alternative pathway	0.3153	3.17	0.0128
NM_015387	NM_015387	preimplantation protein 3		0.3153	3.17	0.0083
AF072928	NM_004685	myotubularin related protein 6	protein dephosphorylation	0.3153	3.17	0.0007
NM_001750	NM_001750; NM_173060; NM_173061; NM_173062	calpastatin isoform a; calpastatin isoform b; calpastatin isoform c; calpastatin isoform d		0.3154	3.17	0.0170
NM_000521	NM_000521	hexosaminidase B preproprotein	glycosphingolipid metabolism; carbohydrate metabolism	0.3154	3.17	0.0003
NM_032146	NM_032146; NM_177976	ADP-ribosylation factor-like 6	small GTPase mediated signal transduction	0.3155	3.17	0.0117
AK054588				0.3156	3.17	0.0159
NM_005509	NM_005509	Dmx-like 1	proteolysis and peptidolysis	0.3158	3.17	0.0001
AK026295				0.3159	3.17	0.0119
NM_002959	NM_002959	sortilin 1 preproprotein	non-selective vesicle targeting	0.3163	3.16	0.0103
AB037797	NM_020801	KIAA1376 protein	sensory perception; signal transduction	0.3163	3.16	0.0029
AB033767	NM_020531	chromosome 20 open reading frame 3	biological_process unknown; biosynthesis	0.3164	3.16	0.0083
NM_006622	NM_006622	serum-inducible kinase	cell cycle; protein amino acid phosphorylation	0.3166	3.16	0.0221
NM_001875	NM_001875	carbamoyl-phosphate synthetase 1, mitochondrial	urea cycle; arginine biosynthesis; pyrimidine base biosynthesis	0.3166	3.16	0.0135
AL157421				0.3167	3.16	0.0091
NM_014109	NM_014109	PRO2000 protein		0.3169	3.16	0.0112
NM_014367	NM_014367	growth and transformation-dependent protein		0.3169	3.16	0.0000
NM_014398	NM_014398	lysosomal-associated membrane protein 3	cell proliferation	0.3169	3.16	0.0022

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_005367	NM_005367	melanoma antigen, family A, 12	biological_process unknown	0.3170	3.15	0.0037
NM_013241	NM_013241	formin homology 2 domain containing 1		0.3170	3.15	0.0022
NM_014714	NM_014714	KIAA0590 gene product		0.3172	3.15	0.0092
NM_022371	NM_022371	torsin family 3, member A		0.3174	3.15	0.0116
BC002924		MGC1842 protein		0.3174	3.15	0.0189
NM_002765	NM_002765	phosphoribosyl pyrophosphate synthetase 2	ribonucleoside monophosphate biosynthesis; nucleoside metabolism; nucleotide biosynthesis	0.3174	3.15	0.0192
NM_022756	NM_022756	hypothetical protein FLJ11730		0.3174	3.15	0.0096
NM_002305	NM_002305	beta-galactosidase binding lectin precursor	apoptosis; heterophilic cell adhesion	0.3176	3.15	0.0091
NM_014763	NM_014763	mitochondrial ribosomal protein L19	protein biosynthesis	0.3177	3.15	0.0071
AF070647				0.3177	3.15	0.0106
NM_016316	NM_016316	REV1-like	DNA repair	0.3178	3.15	0.0064
BC004215		eukaryotic translation elongation factor 1 gamma		0.3179	3.15	0.0239
AF146760	NM_144710; NM_178584	septin 10 isoform 1; septin 10 isoform 2		0.3181	3.14	0.0402
NM_022753	NM_022753	hypothetical protein FLJ12903		0.3182	3.14	0.0153
NM_001530	NM_001530; NM_181054	hypoxia-inducible factor 1, alpha subunit isoform 1; hypoxia-inducible factor 1, alpha subunit isoform 2	response to stress; regulation of transcription, DNA-dependent; signal transduction	0.3182	3.14	0.0121
NM_014504	NM_014504	RAB guanine nucleotide exchange factor (GEF) 1		0.3182	3.14	0.0224
NM_014765	NM_014765	translocase of outer mitochondrial membrane 20 homolog	mitochondrial translocation; protein targeting	0.3185	3.14	0.0127
NM_021953	NM_021953	forkhead box M1	regulation of transcription, DNA-dependent; response to oxidative stress; transcription from Pol II promoter	0.3188	3.14	0.0005
NM_003851	NM_003851	cellular repressor of E1A-stimulated genes	development; cell proliferation; regulation of transcription from Pol II promoter	0.3188	3.14	0.0000
NM_014302	NM_014302	Sec61 gamma	protein targeting	0.3188	3.14	0.0090
NM_032874	NM_032874	hypothetical protein MGC15438		0.3188	3.14	0.0392
NM_005035	NM_005035	mitochondrial DNA-directed RNA polymerase precursor	DNA replication, priming; transcription	0.3189	3.14	0.0068
AL049714				0.3191	3.13	0.0179
NM_016098	NM_016098	brain protein 44-like		0.3192	3.13	0.0063
NM_016044	NM_016044	CGI-105 protein	metabolism	0.3192	3.13	0.0263
NM_002296	NM_002296; NM_194442	lamin B receptor		0.3192	3.13	0.0086
NM_005389	NM_005389	protein-L-isoaspartate (D-aspartate) O-methyltransferase	protein amino acid methylation; protein modification	0.3195	3.13	0.0208
NM_012207	NM_012207; NM_021644	heterogeneous nuclear ribonucleoprotein H3 isoform a; heterogeneous nuclear ribonucleoprotein H3 isoform b	mRNA processing	0.3197	3.13	0.0008
NM_030766	NM_030766; NM_138722; NM_138723; NM_138724	apoptosis regulator BCL-G isoform 2; apoptosis regulator BCL-G isoform 1; apoptosis regulator BCL-G isoform 3		0.3197	3.13	0.0128
NM_021141	NM_021141	ATP-dependant DNA helicase II	double-strand break repair via nonhomologous end-joining; DNA	0.3198	3.13	0.0032

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
			recombination			
NM_003176	NM_003176	synaptonemal complex protein 1	synapsis; meiosis; meiotic recombination; spermatogenesis; cytokinesis	0.3199	3.13	0.0051
NM_014810	NM_014810	centrosome-associated protein 350		0.3199	3.13	0.0153
NM_003363	NM_003363	ubiquitin specific protease, proto-oncogene	ubiquitin cycle; cell growth and/or maintenance; ubiquitin-dependent protein catabolism	0.3199	3.13	0.0159
NM_003416	NM_003416	zinc finger protein 7 (KOX 4, clone HF.16)	regulation of transcription, DNA-dependent; development	0.3201	3.12	0.0050
NM_006070	NM_006070	TRK-fused gene		0.3202	3.12	0.0017
NM_001417	NM_001417	eukaryotic translation initiation factor 4B	regulation of translational initiation; protein biosynthesis	0.3202	3.12	0.0100
AF332223	NM_139214	TGFB-induced factor 2-like, Y-linked	regulation of transcription, DNA-dependent	0.3208	3.12	0.0060
AJ006835				0.3210	3.11	0.0040
NM_024997	NM_024997	activating transcription factor 7 interacting protein 2		0.3211	3.11	0.0007
NM_016029	NM_016029	dehydrogenase/reductase (SDR family) member 7	metabolism	0.3212	3.11	0.0003
NM_003875	NM_003875	guanine monophosphate synthetase	purine nucleotide biosynthesis; purine base biosynthesis; GMP biosynthesis; glutamine metabolism	0.3212	3.11	0.0137
NM_021824	NM_021824	NIF3 NGG1 interacting factor 3-like 1		0.3215	3.11	0.0044
NM_012447	NM_012447; NM_024070	stromal antigen 3	synaptonemal complex formation	0.3215	3.11	0.0241
NM_032413	NM_032413; NM_197955	normal mucosa of esophagus specific 1		0.3216	3.11	0.0134
NM_016467	NM_016467	hypothetical protein LOC51240		0.3218	3.11	0.0009
L38951	NM_002265	karyopherin beta 1	protein-nucleus import, docking; NLS-bearing substrate-nucleus import; protein-nucleus import, translocation; intracellular protein transport	0.3218	3.11	0.0402
NM_001871	NM_001871	pancreatic carboxypeptidase B1 precursor	proteolysis and peptidolysis	0.3219	3.11	0.0096
NM_012215	NM_012215	meningioma expressed antigen 5 (hyaluronidase)	glycoprotein catabolism	0.3221	3.10	0.0098
NM_024824	NM_024824	nuclear protein UKp68		0.3223	3.10	0.0281
AL049998				0.3226	3.10	0.0066
NM_020347	NM_020347	leucine zipper transcription factor-like 1		0.3227	3.10	0.0040
NM_004046	NM_004046	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle	proton transport; ATP biosynthesis	0.3227	3.10	0.0194
NM_005530	NM_005530	isocitrate dehydrogenase 3 (NAD+) alpha precursor	tricarboxylic acid cycle; carbohydrate metabolism	0.3228	3.10	0.0051
NM_003799	NM_003799	RNA (guanine-7-) methyltransferase	mRNA capping	0.3229	3.10	0.0102
NM_003625	NM_003625	PTPRF interacting protein alpha 2	cell-matrix adhesion	0.3229	3.10	0.0214
NM_032487	NM_032487	actin related protein M1		0.3230	3.10	0.0144
NM_001186	NM_001186	BTB and CNC homology 1, basic leucine zipper transcription factor 1	regulation of transcription, DNA-dependent	0.3232	3.09	0.0047

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
BC013128	NM_001806	CCAAT/enhancer binding protein gamma	regulation of transcription, DNA-dependent	0.3232	3.09	0.0007
NM_024678	NM_024678	hypothetical protein FLJ23441	aspartyl-tRNA aminoacylation	0.3233	3.09	0.0002
AK000003	NM_015889	positive cofactor 2, glutamine/Q-rich-associated protein	neurogenesis; transcription from Pol II promoter	0.3233	3.09	0.0006
NM_024740	NM_024740	disrupted in bipolar disorder 1		0.3235	3.09	0.0010
AK055653				0.3236	3.09	0.0016
BC017998	NM_145658	sperm equatorial segment protein 1		0.3239	3.09	0.0271
NM_000983	NM_000983	ribosomal protein L22 proprotein	protein biosynthesis	0.3241	3.09	0.0169
NM_006931	NM_006931	solute carrier family 2 (facilitated glucose transporter), member 3	glucose transport; carbohydrate metabolism; carbohydrate transport	0.3241	3.09	0.0052
NM_005566	NM_005566	lactate dehydrogenase A	glycolysis	0.3241	3.09	0.0084
AK001452				0.3243	3.08	0.0019
NM_015681	NM_015681	B9 protein		0.3243	3.08	0.0089
NM_004988	NM_004988	melanoma antigen, family A, 1		0.3243	3.08	0.0019
NM_003103	NM_003103; NM_032195; NM_058183; NM_138925; NM_138926; NM_138927	SON DNA-binding protein isoform G; SON DNA-binding protein isoform B; SON DNA-binding protein isoform E; SON DNA-binding protein isoform A; SON DNA-binding protein isoform C; SON DNA-binding protein isoform F	anti-apoptosis; biological_process unknown	0.3244	3.08	0.0196
NM_003905	NM_003905	amyloid beta precursor protein-binding protein 1	signal transduction	0.3244	3.08	0.0095
NM_007216	NM_007216; NM_181507; NM_181508	Hermansky-Pudlak syndrome 5 isoform b; Hermansky-Pudlak syndrome 5 isoform a		0.3244	3.08	0.0078
NM_006398	NM_006398	diubiquitin	proteolysis and peptidolysis; protein modification; antimicrobial humoral response (sensu Vertebrata)	0.3245	3.08	0.0063
NM_016400	NM_016400	Huntingtin interacting protein K		0.3246	3.08	0.0127
AB007952	NM_015176	KIAA0483 protein		0.3246	3.08	0.0121
AL080234				0.3247	3.08	0.0185
AB014566	NM_014992	dishevelled-associated activator of morphogenesis 1		0.3248	3.08	0.0240
AK026873				0.3249	3.08	0.0131
NM_004417	NM_004417	dual specificity phosphatase 1	cell cycle; response to oxidative stress; protein amino acid dephosphorylation	0.3249	3.08	0.0007
M82882	NM_172373	E74-like factor 1 (ets domain transcription factor)		0.3250	3.08	0.0052
NM_001903	NM_001903	catenin (cadherin-associated protein), alpha 1, 102kDa	cell adhesion	0.3250	3.08	0.0143
AK056001	NM_005506	scavenger receptor class B, member 2	cell adhesion	0.3252	3.08	0.0198
NM_052961	NM_052961; NM_138718	solute carrier family 26, member 8 isoform a; solute carrier family 26, member 8 isoform b		0.3253	3.07	0.0004
NM_025078	NM_025078	hypothetical protein FLJ22378		0.3254	3.07	0.0039
NM_014764	NM_014764	DAZ associated protein 2	transport	0.3254	3.07	0.0004
NM_021822	NM_021822	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G		0.3258	3.07	0.0110
NM_024045	NM_024045	nucleolar protein GU2		0.3260	3.07	0.0109
NM_032412	NM_032412	putative nuclear protein ORF1-FL49		0.3260	3.07	0.0095
NM_001512	NM_001512	glutathione S-transferase A4	response to stress; metabolism	0.3264	3.06	0.0040

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_016289	NM_016289	MO25 protein		0.3264	3.06	0.0010
AL035563				0.3266	3.06	0.0008
NM_016397	NM_016397; NM_198976	TH1-like protein		0.3267	3.06	0.0156
AK026669	NM_006519	t-complex-associated-testis-expressed 1-like 1	biological_process unknown	0.3268	3.06	0.0059
NM_004548	NM_004548	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10, 22kDa		0.3268	3.06	0.0147
AL050199	NM_181706	hypothetical protein LOC120526		0.3268	3.06	0.0026
NM_025155	NM_025155	hypothetical protein FLJ11848		0.3270	3.06	0.0091
NM_016015	NM_016015; NM_016309	leucine carboxyl methyltransferase	protein modification	0.3271	3.06	0.0060
NM_005655	NM_005655	TGFB inducible early growth response	negative regulation of cell proliferation; TGFbeta receptor signaling pathway; regulation of transcription, DNA-dependent; skeletal development; cell-cell signaling; negative regulation of transcription from Pol II promoter	0.3273	3.06	0.0103
NM_001288	NM_001288	chloride intracellular channel 1	chloride transport; ion transport	0.3273	3.05	0.0049
NM_020198	NM_020198	GK001 protein		0.3275	3.05	0.0075
NM_012459	NM_012459	translocase of inner mitochondrial membrane 8 homolog B	mitochondrial translocation; protein targeting; hearing	0.3278	3.05	0.0107
NM_006901	NM_006901	myosin IXA	vision	0.3280	3.05	0.0116
NM_002494	NM_002494	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	mitochondrial electron transport, NADH to ubiquinone	0.3280	3.05	0.0255
AB058743	NM_025137	hypothetical protein FLJ21439	carbohydrate metabolism	0.3281	3.05	0.0021
NM_012198	NM_012198	grancalcin, EF-hand calcium binding protein	membrane fusion	0.3281	3.05	0.0209
NM_001179	NM_001179	ADP-ribosyltransferase 3	protein amino acid ADP-ribosylation	0.3282	3.05	0.0278
AL117621				0.3282	3.05	0.0085
NM_014320	NM_014320	heme binding protein 2		0.3283	3.05	0.0066
NM_004301	NM_004301; NM_177989; NM_178042	BAF53a isoform 1; BAF53a isoform 2	response to pest/pathogen/parasite; chromatin remodeling; signal transduction	0.3284	3.05	0.0139
NM_031289	NM_031289; NM_153823	germ cell associated 1		0.3285	3.04	0.0158
NM_003379	NM_003379	villin 2	cytoskeletal anchoring	0.3286	3.04	0.0088
AL049331				0.3286	3.04	0.0023
AL049974				0.3286	3.04	0.0007
NM_000463	NM_000463	UDP glycosyltransferase 1 family, polypeptide A1	digestion; estrogen metabolism; bilirubin conjugation	0.3286	3.04	0.0045
NM_017737	NM_017737	hypothetical protein FLJ20275	signal transduction	0.3288	3.04	0.0137
AB051544	NM_018187	KIAA1757 protein		0.3289	3.04	0.0182
AL050372	NM_001441	fatty acid amide hydrolase	fatty acid metabolism	0.3290	3.04	0.0072
NM_005249	NM_005249	forkhead box G1B	brain development; transcription regulation; embryogenesis and morphogenesis	0.3290	3.04	0.0134
AF022789	NM_182488	ubiquitin-specific protease 12-like 1	ubiquitin-dependent protein catabolism; protein modification	0.3290	3.04	0.0176
NM_052888	NM_052888	KIAA0563-related gene		0.3293	3.04	0.0071
AK025435				0.3293	3.04	0.0026
AK021437				0.3294	3.04	0.0193

Systematic ID	RefSeq	Product	GO biological process	Normalized	Ratio	P-value
NM_018848	NM_018848; NM_170784	McKusick-Kaufman syndrome protein	gonad development; heart development; protein folding	0.3294	3.04	0.0195
NM_007106	NM_007106	ubiquitin-like 3		0.3296	3.03	0.0054
NM_003151	NM_003151	signal transducer and activator of transcription 4	JAK-STAT cascade; intracellular signaling cascade; regulation of transcription from Pol II promoter	0.3296	3.03	0.0012
AK027306	NM_030782	cisplatin resistance related protein CRR9p		0.3298	3.03	0.0019
NM_005135	NM_005135	potassium chloride cotransporter KCC3	chloride transport; amino acid transport; sodium ion transport; potassium ion transport	0.3298	3.03	0.0046
NM_032918	NM_032918	RAS-like, estrogen-regulated, growth-inhibitor	small GTPase mediated signal transduction	0.3298	3.03	0.0033
AF065241		thioredoxin delta 3		0.3299	3.03	0.0177

VITA

David Chase Cameron Sprague

Current Position: Medical Scientist
Scott & White
Department of Obstetrics and Gynecology
Office Phone: 254-724-0106
E-mail Address: dsprague@swmail.sw.org

Personal Information: Home Address: 3108 Oakridge Drive, Temple, TX
Marital status: Married to Dawn Carolyn Lux Sprague
Children: William Wyatt Sprague
Maren Elizabeth Sprague

Educational Background:

Year	Degree	Major Area
1995	B.S. (Magna Cum Laude), St. Edward's University	Biology
2006	Ph.D., Texas A&M University	Medical Sciences

Other Postgraduate Education:
May-July, 1999: Participated in the six-week Frontiers in Reproduction Course at Woods Hole Marine Biological Laboratory.

Positions and Appointments:

Scott & White, Temple, Texas, 1996-present
Medical Scientist, Department of Obstetrics and Gynecology
Research Technologist and Laboratory Manager, Core Laboratory for Applied Molecular Biology and Genetics, 1996-2006

Saint Edward's University, Austin, Texas, 1991-1995
Teaching Assistant, Dr. Jimmy T. Mills' Microbiology Laboratory, Spring 1995
Teaching Assistant, Dr. Jimmy T. Mills' Botany Laboratory, Fall 1994